Math

CANADIAN JOURNAL OF MATHEMATICS

Journal Canadien de Mathématiques

VOL. XIII - NO. 1
1961

Published for

THE CANADIAN MATHEMATICAL CONGRESS

by the

University of Toronto Press

The chief languages of the *Journal* are English and French.

Manuscripts for publication in the *Journal* should be sent to the *Editor-in-Chief*, G. F. D. Duff, University of Toronto. Authors are asked to write with a sense of perspective and as clearly as possible, especially in the introduction. Regarding typographical conventions, attention is drawn to the Author's Manual of which a copy will be furnished on request.

All other correspondence should be addressed to the *Managing Editor*, G. de B. Robinson, University of Toronto.

The *Journal* is published quarterly. Subscriptions should be sent to the *Managing Editor*. The price per volume of four numbers is \$10.00. This is reduced to \$5.00 for individual members of recognized Mathematical Societies.

The Canadian Mathematical Congress gratefully acknowledges the assistance of the following towards the cost of publishing this Journal:

University of Alberta
University of British Columbia
Dalhousie University
Université Laval
University of Manitoba
McMaster University
Mount Allison University
Queen's University
University of Saskatchewan
Assumption University
Carleton College
Ecole Polytechnique
Loyola College
McGill University
Université de Montréal
Nova Scotia Technical College
St. Mary's University
University of Toronto

National Research Council of Canada
and the
American Mathematical Society

AUTHORIZED AS SECOND CLASS MAIL, POST OFFICE DEPARTMENT, OTTAWA

y,
r,

e
re
e,
s,
pe

eg

at
rs
d

ne
l:

e

rat
suc
He
are
(2
to
ave
res
$\phi(x$
sat
the
dxa
wit
way
fiel
of
tha
The
tion
but
rela
in F
T
and
that
valu
dete
to s

Re
Rese
autho
the U

14
605
8818

FOUNDATIONS OF THE THEORY OF DYNAMICAL SYSTEMS OF INFINITELY MANY DEGREES OF FREEDOM, II

I. E. SEGAL

1. Introduction. The notion of quantum field remains at this time still rather elusive from a rigorous standpoint. In conventional physical theory such a field is defined in essentially the same way as in the original work of Heisenberg and Pauli **(1)** by a function $\phi(x, y, z, t)$ on space-time whose values are operators. It was recognized very early, however, by Bohr and Rosenfeld **(2)** that, even in the case of a free field, no physical meaning could be attached to the values of the field at a particular point—only the suitably smoothed averages over finite space-time regions had such a meaning. This physical result has a mathematical counterpart in the impossibility of formulating $\phi(x, y, z, t)$ as a bona fide operator for even the simplest fields (in any fashion satisfying the most elementary non-trivial theoretical desiderata), while on the other hand for suitable functions f, the integral $\int\phi(x, y, z, t)f(x, y, z, t)\, dxdydzdt$ could be so formulated. This mathematical development began with the work of Fock **(3)**, in which the field was treated in the conventional way without smoothing, but which gave a concrete representation for a free field that was capable of extension to a representation by bona fide operators, of the smoothed field operators, in the non-relativistic case, an observation that formed the basis for the independent work of Friedrichs **(4)** and Cook **(5)**. The latter gave in rigorous terms the basic mathematical theory of the situation. Additional complications arise in giving an effective relativistic treatment, but it is now established that the suitably smoothed averages of the standard relativistic free real fields may be formulated as bona fide self-adjoint operators in Hilbert space in the strict mathematical sense (see below).

There has not yet been analogous progress for the case of interacting fields, and in the work of Wightman **(6)**, for example, it has merely been postulated that field averages could be given meaning as operators. The expectation values of functions of these operators in the so-called physical vacuum state determine the observable consequences of the theory, and instead of attempting to specify the theory by partial differential equations one may rather attempt

Received August 31, 1959. This research was supported in part by the Office of Scientific Research, and also at an earlier stage by the National Science Foundation, at which time the author enjoyed the hospitality of the Institutes of Mathematics and Theoretical Physics of the University of Copenhagen.

this through a more direct description of such vacuum expectation values. The efforts of Källén and Wightman (**7**) have been directed towards a determination of possible forms for the vacuum expectation values of simple products of smoothed field operators, and in the case of the triple product certain highly refined analytical information has been obtained from the postulates of Lorentz-invariance, microcausality (= "local commutativity"), and positivity of the energy. The finiteness of the vacuum expectation values of such products may, however, be questioned, and in fact in conventional unrenormalized field theory they appear as infinite. In addition, even in the hypothetical case that these values are all finite, they do not necessarily fix the theory, that is, the vacuum expectation values of *all* (smooth) functions of the field operators.

The situation is in a way rather similar to, although vastly more complex than, that with regard to the specification of a probability distribution by its moments. The moments need not be finite; and even when they are finite they do not necessarily determine the distribution (cf. for example (**8**)). The argument that such expectation values must be finite because they have simple physical interpretations (**9A**) is quite parallel to the argument that the second moment of the distribution on the line with element of probability $\pi^{-1}(\alpha^2 + x^2)^{-1}dx$ must be finite because it measures the physical parameter α representing the dispersion of the distribution; speaking loosely in the manner of conventional physical theory, this second moment is easily seen to be proportional to α by an "infinite constant."

A rather natural way to attempt to remedy this situation is to pursue the analogue in the field-theoretic case for the characteristic function in the theory of probability distributions. This is always finite, is known to determine the distribution, and moreover is capable of being characterized intrinsically. The present paper obtains such an analogue in connection with a general (non-pathological) state of a linear field. An interacting field on a particular spacelike surface can be transformed into a linear field (by taking it in the so-called interaction representation, as described for example in (**9B**)), whereupon its vacuum state transforms into an (analytically rather inaccessible) state of the linear field. The present results thereby have implications for the vacuum state of an interacting field. It would in certain respects be more useful to be able to treat the interacting fields directly, but the mathematically ambiguous character of such fields at present seems to make it out of the question to give any *rigorous* treatment of the matter, and in addition there is the apparent lack of any *formal* characterization for the generating functions $E[e^{i\int \phi f}]$ (where E is the vacuum state expectation functional, ϕ is an interacting field, and f a general smoothing function) that would form the analogue for the "Heisenberg" fields of the present functional for a linear field.

At any rate, we show here (rigorously) that a *regular state of a linear field* (of arbitrary unitary transformation properties) *can be characterized by a functional on the corresponding classical wave functions.* That is to say, for

example, that any regular state of the *quantized* hermitian Klein-Gordon field satisfying the equation

$$\Box \phi = m^2\phi$$

is determined by a functional defined on the manifold of all *classical* real-valued normalizable solutions of this equation. The generating functionals that arise in this way are characterized intrinsically, and it is shown how the state may be recovered from the functional. The attainment of these results requires the suitable grouping together of a fairly wide array of scientific developments, but the proofs do not involve any individual points of great technical difficulty.

The present generating functional thus appears as considerably more economical, and mathematically distinctly more viable, than the characterization of a state through the expectation values of products of field operators. It has, however, a rather less direct connection with conventional practice in so-called renormalization theory than the product approach.

In the present paper only Bose-Einstein fields are treated, but the same methods can be adapted to the case of Fermi-Dirac fields.

2. The general linear boson field. The conventional treatments of linear field theory start from specific sets of linear partial differential equations, and arrive at formal operator-valued functions satisfying the same partial differential equations and certain non-trivial commutation relations, after a procedure that varies somewhat from equation to equation. The treatment for the photon case is in particular rather parallel to that for the scalar meson case, but involves additional technical complications, which are somewhat space-consuming and significantly complicate the notation. In addition these treatments have no immediate extension to systems that may be defined not by partial differential equations in ordinary space-time, but in a more general space-time manifold; or which are covariant not with respect to the Lorentz group, but with respect to a more general one. A further difficulty is a fundamental lack of uniqueness—for any given linear quantum field, there exist infinitely many others, satisfying the same commutation relations and partial differential equations, but no two of which are connected by a unitary transformation (cf. **(10)**).

It is therefore relevant that there is available a perfectly general, rigorous, and quite mechanical procedure for linear quantization, whenever the states of the classical system being considered form a complex Hilbert space (or in fact, somewhat more generally). The commutation relations in particular are fixed once the structure of this Hilbert space is specified, and no further examination of the field equations is required. This unique mathematical structure may appropriately be called the "general linear boson field"; its use makes it possible to deal with the commutation relations for extensive classes of fields without the burden of complicated singular functions in the formalism, or the need to utilize generalized functions such as Schwartz' distributions in order to rigorize parts of the analysis. The quantization of a

photon field is in particular, for example, reduced to the *classical* (that is, unquantized) problem of showing that the real normalizable solutions of Maxwell's equations in a vacuum form a complex Hilbert space in a unique Lorentz-invariant manner.

More generally, the normalizable classical solutions of a relativistic linear field equation form the complex Hilbert space $\mathfrak{H}$ that is basic in the following for the treatment of the corresponding field. To present this development in the most elementary fashion, consider for example the real solutions of the Klein-Gordon equation

$$\Box\, \phi = m^2\phi.$$

This is to be interpreted as a heuristic equation, for the relevant solutions are not necessarily conventional functions, but generalized ones. For a rigorous treatment it is simplest to take the Fourier transform Φ of ϕ as basic. In such terms $\mathfrak{H}$ consists of all complex-valued Φ on the hyperboloid $k^2 = m^2$ (here k is the vector with components (k_0, k_1, k_2, k_3) and k^2 denotes the Lorentz squared-length, $k^2 = k_0^2 - k_1^2 - k_2^2 - k_3^2$), such that $\Phi(-k) = \overline{\Phi}(k)$ (corresponding to the reality of the field) and with

$$|\Phi|^2 = \int |\Phi(k)|^2 d\chi(k) < \infty,$$

where $d\chi(k) = |k_0|^{-1} dk_1 dk_2 dk_3$, and is characterized as the unique regular measure on the hyperboloid (within a constant factor) that is Lorentz-invariant. The physical significance of the finiteness condition is more apparent if one deals, equivalently, with positive frequency rather than real solutions of the Klein-Gordon equation (cf. §3 of **(11)**), for these are conventionally interpretable as single-particle wave functions, and the normalizability corresponds to the existence in physical principle of an *individual* free particle (a non-normalizable wave function such as a plane wave having a somewhat ambiguous interpretation as a *beam* of particles).

The description of the corresponding linear quantum field involves basically the formulation and labelling of the field observables, and in particular the specification of the commutation relations of the field variables. Conventionally this is achieved in a heuristic fashion, by postulating the existence of an essentially unique (that is, unique within unitary equivalence, when irreducibility is present) operator-valued function ϕ such that

$$\Box\, \phi = m^2\phi$$

and the commutator $[\phi(x), \phi(x')] = -iD(x - x')$, where D is a certain singular function, and x is written in place of the 4-tuple (x, y, z, t). No such operator-valued function is known to exist, and actually there is practically conclusive evidence that it cannot exist in any literal sense; and in any event it could not be unique within unitary equivalence even if its range of values formed an irreducible set of operators.

In order to deal in a mathematically clear and physically conservative way with such a matter, it is appropriate to make first a purely mathematical

construction and development; next to make a statement of what *mathematical* objects in this construction represent the observable or theoretically relevant *physical* objects associated with the physical system motivating the construction; and finally to establish the essential agreement between the resulting physical theory and the conventional one and/or experimental indications. The first part of this procedure should be mathematically rigorous; the second part should be precise, but necessarily only in the sense of legal rather than mathematical definitions; while the third part may well involve quite heuristic elements, and in fact this is necessarily the case when dealing with a theory whose conventional form is heuristic, as is quantum field theory. Accordingly we proceed as follows.

Mathematical construction. There is no compelling reason not to use terminology here that is indicative of the physical object being considered, and a great deal of circumlocution may be avoided in this way. In particular, the term "field" will be so used, but several different mathematical objects related to various heuristic types of fields must be distinguished. It will suffice here to deal with *concrete*, *general*, *clothed*, and *zero-interaction* linear boson fields.

Definition 1. Let $\mathfrak{H}$ be a real linear vector space, and let B be a given skew-symmetric bilinear form over $\mathfrak{H}$.

(a) A *concrete* LBF (LBF = linear boson field) over $(\mathfrak{H}, B)$ is a map $z \rightarrow R(z)$ from to $\mathfrak{H}$ the self-adjoint operators in a complex Hilbert space $\mathfrak{K}$ such that

$$\text{(x)} \qquad e^{iR(z)}e^{iR(z')} = e^{iR(z+z')}e^{iB(z,z')} \qquad (z, z' \text{ arbitrary in } \mathfrak{H}).$$

Two such fields, $R(.)$ on $\mathfrak{K}$ and $R'(.)$ on $\mathfrak{K}'$, over the same $(\mathfrak{H}, B)$, are *unitarily equivalent* in case there exists a unitary operator V from $\mathfrak{K}$ onto $\mathfrak{K}'$ such that

$$VR(z)V^{-1} = R'(z), \text{ for all } z \text{ in } \mathfrak{H}.$$

(b) A *bounded field observable* of a given concrete LBF is defined as a bounded operator on $\mathfrak{K}$ that is a limit of a uniformly convergent sequence of operators, each of which is in the weakly closed ring of operators generated by the $e^{iR(z)}$, for z ranging over some finite-dimensional (but otherwise arbitrary) linear subspace of $\mathfrak{H}$. The set of all bounded field observables is then a uniformly closed self-adjoint algebra of operators on $\mathfrak{K}$ (cf. **(12)**, referred to henceforth as "I").

(c) Two concrete LBF's $R(.)$ and $R(.)'$ on $\mathfrak{K}$ and $\mathfrak{K}'$, over the same $(\mathfrak{H}, B)$, are said to be *physically equivalent* in case there is a one-to-one correspondence between their respective bounded hermitian field observables preserving the operations of addition and squaring. A general LBF over $(\mathfrak{H}, B)$ is defined as a physical-equivalence class of concrete LBFs. We note that when $\mathfrak{H}$ is a complex Hilbert space and B is the canonically associated skew form (cf. below), there is only one general LBF.

(d) A *clothed* LBF over $(\mathfrak{H}, B)$ is a couple consisting of a general LBF over $(\mathfrak{H}, B)$, together with a given state E of the (abstract) algebra of field observables. E is said to be *regular* in case its restriction to the weakly closed ring of operators generated by the $e^{iR(z)}$, as z ranges over a finite-dimensional subspace of $\mathfrak{H}$ on which B is non-degenerate, is weakly continuous relative to the unit sphere of this ring, for every such finite-dimensional subspace, and some concrete representative for the general LBF. An LBF is *properly clothed* if the associated state E is regular.

(f) When a real linear vector space $\mathfrak{H}$ has in addition a designated structure as a complex Hilbert space, compatible with its real-linear structure, the notation $(\mathfrak{H}, B)$ will be understood to refer to $\mathfrak{H}$ as a real linear vector space, with $B(z, z') = \frac{1}{2}Im[(z, z')]$; and the notation $\mathfrak{H}$ alone may refer to the couple $(\mathfrak{H}, B)$, when it is clear from the context that it is this couple that is relevant.

(g) As an example of a clothed LBF the *zero-interaction* LBF is defined as that clothed LBF over the complex Hilbert space $\mathfrak{H}$ for which the given state E is invariant under the induced action of all unitary operators on $\mathfrak{H}$ (cf. I, and especially Cor. 3.1 showing the uniqueness of the free LBF).

Physical interpretation. If ϕ is for example the conventional real Klein-Gordon field and f is any smooth function on space-time that vanishes at infinity, the field average $\int \phi(x) f(x) d_4x$ is just such an $R(z)$. The appropriate $\mathfrak{H}$ is just that defined above, consisting of all normalizable solutions of the Klein-Gordon equation; the appropriate z is that function on the mass hyperboloid (that is, the manifold $k^2 = m^2$) that coincides there with the complex conjugate of the Fourier transform of f; and the appropriate skew-symmetric form $B(z, z')$ is just that defined equivalently as $B(z, z') = \frac{1}{2}\iint D(x - x') f(x) f'(x') d_4x d_4x'$, where z' is related to f' in the same fashion as z to f, and D is the conventional singular function such that $[\phi(x), \phi(x')] = -iD(x - x')$, or, in a form in which the finiteness of $B(z, z')$ is more apparent,

$$B(z, z') = -\frac{i}{8\pi^3} \int_M \operatorname{sgn} k_0 \, z'(k) z(-k) d\lambda(k)$$

where M denotes the mass hyperboloid. The equation (x) is the bounded (Weyl) form of the infinitesimal relation

$$[R(z), R(z')] = -2iB(z, z'),$$

to which it is formally equivalent.

Conventionally it was assumed, then, that the Klein-Gordon field is concrete and irreducible, and that this sufficed to define the field uniquely. The discovery that in actuality this was very far from being the case led to the introduction of general linear fields in I, in which the more sophisticated form of quantum phenomenology developed originally in **(13)** is employed. The notion of physical equivalence above seems at first glance insufficiently restrictive, but it is shown in **(13)** that it implies that the two systems have

corresponding pure states, corresponding observables have identical spectral values and probability distributions in given states, etc., so that the two systems are in fact in all observable respects the same.

The distinguished state of a clothed LBF is physically the vacuum state, and a more conventional formulation may be obtained from the correspondence between states and representations of operator algebras, which leads to the result (cf. I) that for any properly clothed LBF there is a concrete LBF with a distinguished vector v, whose transforms under the $e^{iR(z)}$ span the representation space $\mathfrak{K}$, and plays the role of the conventional physical vacuum state vector, in that the vacuum expectation value of the observable represented by the operator A is (Av, v). Conversely, such a concrete LBF with distinguished vector v gives rise to a properly clothed LBF associated with it in the foregoing fashion; and the concrete LBF-with-vector is uniquely determined, within unitary equivalence, by the clothed LBF.

The definition of regular state is rather technical, but is admissible from a purely physical point of view, since only those values of the state on the field observables are involved; and is justified by the existence of various equivalent formulations. It is surely reasonable from an empirical-physical viewpoint to require that for a physical state E, $E[e^{itR(z)}]$ be a continuous function of t, for any fixed vector z in $\mathfrak{H}$, and the regular states are precisely those that have this property, and in addition are determined in a natural way by the expectation values $E[e^{iR(z)}]$ for all z. Alternatively, a regular state is one whose restriction to any subsystem of a finite number of degrees of freedom (that is, the ring of operators generated by the $e^{iR(z)}$ as z ranges over a finite-dimensional subspace of $\mathfrak{H}$) is a normalizable (possibly mixed) state in essentially the conventional sense, that is, it has the form $E(A) = \mathrm{tr}\,(AD)$ for some operator D of absolutely convergent trace. It should be noted that the present notion of regularity is more stringent than that employed in I, which permitted a theoretical generality that is physically not entirely appropriate. In fact Corollary 3.1 of I is correct (at least in proof) only with the present notion of regular state, by virtue of the possible existence of a pathological state other than the zero-interaction vacuum, which would agree with the zero-interaction vacuum on all sufficiently smooth observables, in particular on all observables that are uniform limits of products of those of the form $f(R(z))$, for some z and continuous function f that vanishes at infinity, but not on the weak limits of such. A state that is not determined by its values on such observables could be fully determined only through the use of infinite fields; it could never be obtained as a limit of states in cut-off theories. The restriction to regular states thus amounts to a type of universally covariant cut-off; in place of it one could substantially limit the observables to those "smooth" ones obtainable in the fashion indicated, which it can reasonably be argued are the only ones that can actually be observed even conceptually.

An interacting relativistic field on a particular space-like surface, in the interaction representation, gives a formal example of a linear boson field clothed

by the physical vacuum; in conventional theory, this clothing degenerates as the space-like surface recedes or advances into the infinite past or future, and the *zero-interaction* LBF is obtained.

Formal equivalence of the present and the conventional formalisms. It must be shown how to define the conventional quantized field $\phi(x)$; that this satisfies the relevant partial differential equation and also the canonical commutation relations. To this end let δ_x' denote the projection of the delta-function at the point x on the manifold of solutions of the relevant partial differential equation, for example, δ_x' is the reciprocal Fourier transform of the function in momentum space that agrees with $e^{ik\cdot x}$ on the mass hyperboloid and vanishes outside the hyperboloid, in the case of the Klein-Gordon equation. Then set $\phi(x) = R(\delta_x')$; δ_x' is an improper element of $\mathfrak{H}$, but this is inevitable since $\phi(x)$ is an improper operator. That ϕ as thus defined satisfies the given partial differential equations is a simple deduction from the Parseval formula for Fourier transforms. That the canonical commutation relations hold follows by substitution of δ_x' and $\delta_{x'}'$ for z and z' in the relation $[R(z), R(z')] = -2iB(z, z')$.

It must also be shown that conversely, from such a conventional quantized field ϕ, the present operators $R(z)$ can be constructed. If f is any smooth function on space-time that vanishes at infinity, the operator $\int \phi(x)f(x)d_4x$ is defined as $R(z)$, with z equal to the projection of f on the space of solutions of the relevant partial differential equation. For a non-scalar field this definition extends with the use of the Lorentz-invariant inner product in the finite-dimensional spin space for the field in question. Although the projection in question is singular as an operator in a Hilbert space, the z's may be analytically well defined for appropriate f, and thereby the $R(z)$ also.

3. Characterization and uniqueness of the generating functional. For any state E of a general linear boson field over $(\mathfrak{H}, B)$, the functional $\mu(z) = E[e^{iR(z)}]$ is well defined, and may be called the *generating functional* for the state. From it, the expectation values of arbitrary products of the field (at distinct points) may be obtained by differentiation, at least heuristically, when such expectation values are finite. For foundational purposes what is essential is

THEOREM 1. *A (complex-valued) function μ on $\mathfrak{H}$ is the generating functional of a regular state E of the general linear boson field over $(\mathfrak{H}, B)$ with non-degenerate B if and only if the restrictions of μ to arbitrary finite-dimensional subspaces of $\mathfrak{H}$ are continuous, $\mu(0) = 1$, and*

$$\sum_{j,k \in F} \mu(z_j - z_k)e^{iB(z_j, z_k)}\bar{\alpha}_k\alpha_j \geq 0$$

for arbitrary z_j in $\mathfrak{H}$ and complex numbers α_j, F being any finite index set. The functional μ uniquely determines E.

The "only if" part follows from the obvious fact that

$$\left(\sum_j \alpha_j e^{iR(z_j)}\right)^*\left(\sum_j \alpha_j e^{iR(z_j)}\right) \geq 0,$$

together with the relation (x). To prove the "if" part, let μ be given satisfying the stated conditions. Put K_0 for the set of all complex-valued functions on $\mathfrak{H}$ that vanish except at a finite set of points, with the inner product

$$(f, g) = \sum_{z, z'} f(z)\bar{g}(z')\mu(z - z')e^{iB(z,z')}$$

(f and g in $\mathfrak{K}_0$). This inner product is a positive semi-definite hermitian form, so the set of all f in $\mathfrak{K}_0$ with $(f, f) = 0$ forms a linear subspace $\mathfrak{K}_0'$ of $\mathfrak{K}_0$, and the quotient $\mathfrak{K}_0/\mathfrak{K}_0' = \mathfrak{K}_1$ (say) has canonically defined on it a strictly positive definite hermitian form:

$$(f', g') = (f, g);\ f' = f + \mathfrak{K}_0' \text{ and } g' = g + \mathfrak{K}_0'.$$

Now let $U_0(z')$, for z' in $\mathfrak{H}$, denote the transformation on $\mathfrak{K}_0$:

$$f(z) \to e^{iB(z',z)}f(z - z').$$

Then $U_0(z')$ is a linear operator with the inverse $U_0(-z')$. Also,

$$(U_0(z')f, U_0(z')g) = (f, g)$$

for arbitrary f, g, and z'. It follows that the map

$$U_1(z'): f' \to U_0(z')f + \mathfrak{K}_0$$

is a well-defined linear transformation on $\mathfrak{K}_1$, that it has the inverse $U_1(-z')$, and that

$$(U_1(z')f', U_1(z')g') = (f', g').$$

Hence $U_1(z)$ extends uniquely to a unitary transformation $U(z)$ on the completion $\mathfrak{K}$ of $\mathfrak{K}_1$. Now $U_0(z)U_0(z') = e^{iB(z,z')}U_0(z + z')$, from which it follows readily that

$$U(z)U(z') = e^{iB(z,z')}U(z + z').$$

In particular, $[U(tz): -\infty < t < \infty]$ is a one-parameter group of unitary operators in $\mathfrak{K}$. This one-parameter group is continuous; to show this it suffices, by the density of $\mathfrak{K}_1$ in $\mathfrak{K}$, to show that $(U(tz)f', g')$ is continuous for arbitrary f' and g' in $\mathfrak{K}_1$. But

$$(U(tz)f', g') = \sum_{u,u'} e^{iB(tz,u)}f(u - tz)\bar{g}(u')\mu(u - u')e^{iB(u,u')}.$$

Now if f has the values $\sigma_1, \ldots, \sigma_n$ at $v_1, \ldots, v_n$, respectively, and g has the values $\tau_1, \ldots, \tau_n$ at these points, and both functions vanish elsewhere, this sum is

$$\sum_{j,k} \mu(v_j + tz - v_k)\, e^{iB(tz, v_j+tz)}\, e^{iB(v_j+tz, v_k)}\sigma_j\bar{\tau}_k,$$

which represents a continuous function of t by the assumption on μ. Thus the one-parameter group has a self-adjoint generator $R(z)$, and the $R(z)$ satisfy the relations (x).

If f_0 denotes the function defined by the equations $f_0(0) = 1, f_0(z) = 0$ for $z \neq 0$, then $(f_0', f_0') = 1$, and

$$(U(z)f_0', f_0') = \mu(z).$$

Thus setting

$$E(A) = (Af_0', f_0'),$$

E is regular and has characteristic functional μ.

To show that μ determines E uniquely, let E' be an arbitrary regular state with characteristic functional μ. Then E' is weakly continuous relative to the unit sphere of the ring $\mathfrak{A}_{\mathfrak{M}}$ generated by the $e^{iR(z)}$ for z in $\mathfrak{M}$, for all $\mathfrak{M}$ on which B is non-degenerate, and some concrete LBF. If E is similarly weakly continuous, etc., relative to the same concrete LBF, then the unicity follows from the circumstance that the finite linear combinations of the $e^{iR(z)}$, z in $\mathfrak{M}$, form an algebra $\mathfrak{A}_{0,\mathfrak{M}}$ whose weak closure is $\mathfrak{A}_{\mathfrak{M}}$; this implies that the unit sphere in $\mathfrak{A}_{0,\mathfrak{M}}$ is weakly dense in that of $\mathfrak{A}_{\mathfrak{M}}$, according to a variant of an argument due to von Neumann **(14)** (for full details cf. **(15)**). The assumed weak continuity of E and E' on $\mathfrak{A}_{\mathfrak{M}}$ relative to the unit sphere, together with their agreement on the unit sphere of $\mathfrak{A}_{0,\mathfrak{M}}$, then implies their equality.

To conclude the proof it therefore suffices to show the

LEMMA. *If a state E is weakly continuous on $\mathfrak{A}_{\mathfrak{M}}$ relative to the unit sphere for one concrete* LBF, *then the same is true for all concrete* LBFs.

To prove this, observe that as B is non-degenerate on $\mathfrak{M}$, co-ordinates may be chosen so that in $\mathfrak{M}$, $z = (x_1, \ldots, x_n) \oplus (y_1, \ldots, y_n)$ ($\mathfrak{M}$ being of dimension $2n$), and B has the form $B(z, z') = \sum_k (x_k y_k' - x_k' y_k)$. The relations (x) then imply those on which von Neumann's proof of the uniqueness of the Schrødinger operators is based, so that by his result, $\mathfrak{A}_{\mathfrak{M}}$ is, within multiplicity, and unitary equivalence, the conventional system of bounded observables in quantum mechanics for a particle with a $2n$-dimensional phase space. That is to say, $\mathfrak{A}_{\mathfrak{M}}$ is unitarily equivalent to an n-fold copy, for some finite or infinite cardinal number n, of the ring of operators generated by the $\exp(isq_j)$ and the $\exp(itp_k)$ $(-\infty < s, t < \infty; j, k = 1, 2, \ldots, n)$ in their action on the space $L_2(E_n)$ of all complex-valued square-integrable functions on Euclidean n-space. All that is relevant here is that the weak topology on the unit sphere of an operator ring is easily seen to be independent of the multiplicity n of its representation. Now for any two concrete LBFs, the resulting $\mathfrak{A}_{\mathfrak{M}}$ are

within unitary equivalence multiples of the same ring of operators, and the lemma follows.

This concludes the proof of the theorem, but it also follows and seems worth pointing out explicitly that we have the

COROLLARY. *If $\mathfrak{M}$ is a linear subspace of $\mathfrak{H}$ on which B is non-degenerate, then the ring $\mathfrak{A}_{\mathfrak{M}}$ of all field observables based on M is a factor of type I_∞, and the restriction of E to $\mathfrak{A}_{\mathfrak{M}}$ has the form*

$$E(X) = \operatorname{tr}(XD_{\mathfrak{M}})$$

for some operator $D_{\mathfrak{M}}$ of absolutely convergent trace relative to $\mathfrak{A}_{\mathfrak{M}}$.

This is an immediate consequence of von Neumann's result used as above, together with the known form of the states of the ring of all bounded operators on a Hilbert space that are weakly continuous relative to the unit sphere. For this last result cf. for example (**16**, Theorem 14).

4. Some examples. The *generating functional of the zero-interaction linear boson field* over a (complex) Hilbert space may be computed explicitly as follows. If $C(z)$ denotes the creation operator for a particle with wave function z, as defined for example in (**5**), and $R(z)$ denotes the closure of $(C(z) + C(z)^*)/\sqrt{2}$, while E denotes the *zero-interaction* vacuum state, then the evaluation of $E[e^{iR(z)}]$ reduces to the case when $\mathfrak{H}$ is one-dimensional (see Cor. 3.6 of (**17**)). The representation of $e^{iR(z)}$ in terms of the one-parameter unitary groups generated by the canonical p's and q's in one dimension leads to a familiar type of integral involving the normal distribution, and ultimately to the result

$$\mu(z) = e^{-\frac{1}{4}\|z\|^2}$$

for the *zero-interaction* vacuum generating functional.

This may be used to obtain zero-interaction vacuum expectation values of arbitrary products of field values by noting that formally one has for any state E,

$$E[R(z_1)\dots R(z_n)] = i^{-n}\{\partial^n/\partial t_1\dots\partial t_n\}E[e^{it_1R(z_1)}\dots e^{it_nR(z_n)}]|_{t_1=\dots=t_n=0},$$

while

$$e^{it_1R(z_1)}\dots e^{it_nR(z_n)} = e^{iR(t_1z_1+\dots+t_nz_n)+i\sum_{j<k} t_jt_kB(z_j,z_k)}$$

which has zero-interaction vacuum expectation value

$$\exp[-\tfrac{1}{4}\sum_{j,k} t_jt_k(z_j, z_k)].$$

Of course, in general the indicated derivative as well as the expectation value of the product of field operators will fail to exist. However, it is easy to justify

in the bare field case the foregoing formal equality, and thereby obtain explicit expressions for the bare vacuum expectation values of products of fields. In the simplest non-vanishing case, there results the formula

$$E[R(z_1)R(z_2)] = \tfrac{1}{2}(z_1, z_2)$$

(also easily obtainable directly) having the conventional interpretation that the *zero-interaction* vacuum expectation value of the product of two field values is (1/2) the singular function providing the kernel for the symmetric form giving the Lorentz-invariant inner product, as is well known for special fields.

Turning to general states, if all the regular states could readily be expressed in terms of zero-interaction field quantites, the use of the general LBF might be avoidable. Theoretically this would be rather extraordinary, in view of the general situation in quantum fields, and in fact quite explicit examples can be constructed to show that this is not the case. To show simply *the existence of regular pure states that are not normalizable in the Fock-Cook representation*, that is, not of the form $E(A) = (Av, v)$ for some normalizable vector v, one may proceed as follows. Take the representation of the bare field in terms of the space $L_2(\mathfrak{H}_r, n)$ of square-integrable functionals over a real subspace $\mathfrak{H}_r$ of $\mathfrak{H}$ such that $\mathfrak{H} = \mathfrak{H}_r + i\mathfrak{H}_r$, as given in **(17)**. Let θ denote the automorphism of the algebra of field observables over $\mathfrak{H}$ taking each canonical p into $2p$ and each canonical q into $q/2$ (this exists by Theorem 2 of I), and define E_θ as the transform of the bare vacuum state under the induced action of θ, that is, $E_\theta(A) = E(A^\theta)$, where E denotes the bare vacuum state and $A \rightarrow A^\theta$ the action of θ. Then E_θ is evidently regular and pure, but may be seen to be non-normalizable as follows.

As a basis for an indirect argument, assume that $E_\theta(A) = (Au, u)$, for some u in $L_2(\mathfrak{H}_r)$. Then for arbitrary x in $\mathfrak{H}_r$, using the notation and Theorem 3 of **(17)**, with $c = \tfrac{1}{2}$,

$$E_\theta(e^{iQ(x)}) = \int_{\mathfrak{H}_r} e^{ix\cdot y}|u(y)|^2 dn(y),$$

and since $E_\theta(e^{iQ(x)}) = E(e^{iQ(x)/2})$, which is readily evaluated as

$$e^{-1/8|x|^2},$$

it suffices to show

$$\int_{\mathfrak{H}_r} e^{ix\cdot y}F(y)dn(y) \neq e^{-1/8|x|^2}$$

if F is an arbitrary element of $L_1(\mathfrak{H}_r, n)$. Such an F is an L_1-limit of polynomial functions, so it suffices to show that if $G(.)$ is any polynomial and

$$g(x) = \int e^{ix\cdot y}G(y)dn(y),$$

then

$$\sigma = \sup_x |g(x) - e^{-1/8\|x\|^2}| > \delta,$$

where δ is > 0 and independent of G. But $g(x)$ has the form

$$g(x) = p(x)e^{-\frac{1}{4}\|x\|^2},$$

where p is a polynomial, based, say, on the finite-dimensional subspace $\mathfrak{M}$ of $\mathfrak{H}$. Now if x is in the orthocomplement of $\mathfrak{M}$, $p(x) = p(0)$, whence

$$\sigma > \inf_a \sup_{0<r<\infty} |ae^{-\frac{1}{4}\|x\|^2} - e^{-1/8\|x\|^2}|.$$

Simple calculus leads from this to the bound $\sigma > \frac{1}{4}$.

5. Special dynamics in terms of generating functionals. For the special but significant case of a Hamiltonian that is quadratic in the canonical variables, there is a remarkably simple formulation of the corresponding quantum dynamics. It gives the explicit time development of the generating functional, and hence of the system, in terms of the corresponding classical dynamics.

If a classical *motion with Hamiltonian quadratic in the canonical variables takes*

$$z \to V_t z(-\infty < t < \infty;\ t = \text{time})$$

then the corresponding quantum-mechanical *motion transforms generating functionals as follows:*

$$\mu(z) \to \mu(V_t z).$$

To prove this, note that if the Hamiltonian is quadratic, then the motion in phase space $\mathfrak{H}$ is linear. That is to say, for any fixed t, V_t is a non-singular transformation preserving the fundamental skew form $B(z, z') = \sum_k (p_k' q_k - p_k q_k')$, where $z = (p_1, \ldots, p_n) \oplus (q_1, \ldots, q_n)$, where n may be finite, or there may be infinitely many degrees of freedom, in which case the appropriate modifications in the notation are obvious (cf. **(18)** for the infinitesimal situation in a finite number of dimensions and **(19)** for the global situation in any number of dimensions). Now for any such transformation, the corresponding quantum-mechanical motion may be uniquely given by the condition that it transform $R(z)$ into $R(Tz)$ (cf. I), so that it has precisely the stated effect on the generating functional.

6. Restriction to regular observables instead of regular states. Instead of dealing with a restricted class of states of an extensive system of observables, one may contemplate dealing with all states of a restricted subsystem of observables. A general bounded self-adjoint operator is an observable only in a quite theoretical way; a class of operators slightly closer to actual measurements than those merely generated by the canonical variables would be those expressible explicitly in terms of them. It turns out that for systems of a finite number of degrees of freedom, which may be used to approximate infinite systems, there is a natural way to make this idea effective.

Definition 2. A *regular observable* over a finite-dimensional linear space $\mathfrak{H}$ with distinguished skew-symmetric form B, relative to a concrete LBF over $(\mathfrak{H}, B)$, is one of the form

$$\int_{\mathfrak{H}} e^{iR(z)} f(z)dz$$

or uniformly approximable by such. Here f is integrable over $\mathfrak{H}$, while dz is the element of measure determined by B.

The foregoing integral may be taken in the strong or weak operator topologies (in the sense of **(20)**). The resulting class of operators is unaffected by a change in the measure employed, within absolute continuity. It is not difficult to verify that the notion of regular observable is invariant under physical equivalence, and therefore may be applied to elements of the general LBF over a finite-dimensional space.

THEOREM 2. *The regular observables form a uniformly closed self-adjoint algebra, every state of which extends uniquely to a regular state of the general* LBF *over* $(\mathfrak{H}, B)$, $\mathfrak{H}$ *being finite-dimensional; and every regular state arises in this way.*

That the regular observables form an algebra follows without difficulty from the relations (x), the general properties of the integrals involved, and the Fubini theorem. Now for any state E,

$$\left| E\left[\int_{\mathfrak{H}} e^{iR(z)} f(z)dz\right]\right| < \int_{\mathfrak{H}} |f(z)|dz,$$

so that by the known form of the general continuous linear functional in L_1, these exists a bounded measurable function μ on $\mathfrak{H}$ such that

$$E\left[\int_{\mathfrak{H}} e^{iR(z)} f(z)dz\right] = \int \mu(z) f(z)dz.$$

By the positivity of E, μ satisfies the inequality

$$\iint \mu(z - z')e^{iB(z,z')} f(z)\bar{f}(z')dz\, dz' \geqslant 0.$$

If μ were continuous, this would imply that μ is a generating functional. Now it is well-known that a measurable integrally-positive-definite function on a locally compact group differs from a continuous positive definite function on the group on a null set (in the sense of Haar measure on the group). An argument similar to that involved in the proof of this result yields the corresponding result here, or this result may be derived from the theory of positive definite functions on groups; here we take the latter course, as this is illuminating in certain additional respects.

Let G be the group (cf. **(18)**) of all pairs (z, s) with z in $\mathfrak{H}$ and s real, and the multiplication

$$(z, s) . (z', s') = (z + z', s + s' + B(z, z')).$$

Then G is a Lie group with the obvious manifold structure. Set $\lambda(z, s) = \mu(z)e^{-is}$; then λ is a measurable and integrally positive definite function on G, as is readily verified. By the result cited, it differs on a null set from a continuous positive definite function. Now $\mathfrak{H}$ may be identified with the subset of G consisting of all elements of the form $(z, 0)$, the map $z \to (z, 0)$ being a homeomorphism into. It follows that μ differs on a null set in $\mathfrak{H}$ from a continuous function μ'.

Now μ' will satisfy the positivity condition given in Theorem 1, as follows from the integral positivity condition by a simple approximation argument. Since $(\mu'(z_j - z_k)e^{iB(z_j, z_k)}; j, k = 1, \ldots, n)$ is a positive semi-definite hermitian matrix for arbitrary $z_1, \ldots, z_n$, and μ' cannot vanish identically since $E \neq 0$, $\mu(0)$ must be positive. Setting $\mu'' = \mu'/\mu'(0)$, μ'' satisfies all the conditions of Theorem 1, so there exists a regular state E' of the general LBF, say $\mathfrak{A}$, over $(\mathfrak{H}, B)$, whose generating functional is μ''. Now the unit sphere of the algebra $\mathfrak{R}$ of regular observables is weakly dense in that of $\mathfrak{A}$, by the result cited above, together with the easily established fact that the weak closure of $\mathfrak{R}$ is $\mathfrak{A}$. In particular,

$$\sup_{X \in \mathfrak{A}, |X|<1} E'(X^*X) = \sup_{X \in \mathfrak{R}, |X|<1} E\,(X^*X)$$

which shows, since E' is a state of $\mathfrak{A}$, that the right side of the foregoing equality has the value unity. On the other hand,

$$E'\left[\int_{\mathfrak{H}} e^{iR(z)} f(z)dz\right] = (\mu'(0))^{-1} \int_{\mathfrak{H}} \mu'(z) f(z)dz$$

for arbitrary continuous f vanishing outside a compact set, as the integrals may then be taken in the Riemann sense; and by a simple approximation argument, it results that $E'(X) = (\mu'(0))^{-1}E(X)$ for arbitrary X in $\mathfrak{R}$. Now as E is a state of $\mathfrak{R}$,

$$\sup_{X \in \mathfrak{R}, |X|=1} E(X^*X) = 1,$$

and it follows that $\mu'(0) = 1$.

Thus there exists a regular state E' of the full general LBF extending the given state E of the regular observables. That E' is unique follows from the density of the unit sphere of $\mathfrak{R}$ in that of $\mathfrak{A}$. Conversely, if E' is a given regular state of $\mathfrak{A}$, its restriction to $\mathfrak{R}$ is a positive linear functional which by the argument just given must have unit norm, and so be a state E.

It should be remarked that the connection with the theory of positive definite functions can be misleading, if not utilized with care. For example, when $B = 0$, the condition of Theorem 1 becomes ordinary positive definiteness (apart from the normalization), but the theorem is then irremediably false, as it asserts essentially that an arbitrary positive definite function is the Fourier-Stieltjes transform of an *absolutely continuous* measure. The introduction of a non-degenerate B thus has roughly the qualitative effect of eliminating the possible discontinuous and continuous but singular parts of the associated state.

7. Possible further developments. A number of problems emerge from the foregoing work. Some typical ones of interest are as follows.

1. The *zero-interaction* vacuum is the only regular state of the general LBF over a complex Hilbert space $\mathfrak{H}$ that is invariant under all unitary operators on $\mathfrak{H}$. Presumably it is, more cogently, the only such state invariant under a physically relevant representation of the Lorentz group, say, that associated with a relativistic particle of integral spin. For a particle of positive mass (=minimum proper value of the infinitesimal generator of translations it time), it is clear that there exist no normalizable states in the Fock-Cook representation other than the zero-interaction vacuum that are invariant, but it remains to be proved that this is the case for all regular states. In the vanishing mass case the situation is less clear, and correspondingly more interesting.

2. It seems probable that any symplectic transformation in a complex Hilbert space $\mathfrak{H}$ (that is, a real-linear transformation leaving invariant the imaginary part of the inner product) will effect a transformation of the zero-interaction vacuum of the general LBF over $\mathfrak{H}$ into a state that is not normalizable in the Fock-Cook representation, except when the transformation is unitary; this would generalize the example given above of such a state.*

3. When $\mathfrak{H}$ is finite-dimensional, $\mu(z)$ is essentially the Fourier transform of Wigner's quasi-probability distribution (**(21)**; cf. also **(22)** and the literature cited there, especially the paper by Moyal). Now a classical motion takes a z into a z', while a quantum-mechanical one takes a generating function μ into another μ'. The determination of the precise relation between $\mu'(z)$ and $\mu(z')$, shown above to be identical in the case of a quadratic Hamiltonian, is connected with the problems of interpretation considered by Wigner and later authors. Although it seems fairly clear that no exact result is to be hoped for, even a simple approximate relation between the two functions might well be quite useful.

4. The difficulty forming the basis of the preceding problem also suggests the more extensive question of the extent to which a theory similar to the present one, but covariant under the entire group of classical contact transformations rather than merely the symplectic group, can be set up. In such a theory the analogue to the smoothed field operators $R(z)$ would perhaps be a function $R(Z)$ defined for infinitesimal contact transformations Z, and satisfying in place of the commutation relations involved above, the relations

$$[R(Z), R(Z')] = R([Z, Z']) + \Omega\,(Z, Z'),$$

where Ω denotes the fundamental second-order differential form on $\mathfrak{H}$ (that is, the well-known form $\sum_k dp_k dq_k$ in the case of a finite number of degrees of freedom, while in the infinite case it is an analogous form determined by the field commutators). The main difficulty here is not the presence of the term

**Remark added in proof.* This result has been established in the meantime by David Shale.

$R([Z, Z'])$, which was absent above because Z and Z' were essentially infinitesimal translations in $\mathfrak{H}$ and so had vanishing commutator, but rather the circumstance that the $\Omega\ (Z, Z')$ are not constant numbers, but scalar functions on $\mathfrak{H}$, the multiplications by which naturally do not commute with the Z's.

A possible way around this difficulty is the employment of a suitable analogue to the group G employed above, such as the group whose Lie algebra (that is, associated infinitesimal group) consists of all pairs (Z, f), where Z is an infinitesimal classical contact transformation and f is a function on phase space, with the commutation relations

$$[(Z,f), (Z',f')] = ([Z, Z'], Zf' - Z'f + \Omega\ (Z, Z')).$$

That these define a Lie algebra (that is, notably that the Jacobi conditions hold) follows from the fact that Ω is a closed form, using the expression for the derivative of a form in terms of the form itself, together with brackets of vector fields and the operations of the vector fields on values of the form (cf. **(23)**, § 1). A pair such as (Z, f) may be interpreted as the generator of a contact transformation in the tangent bundle of the phase space $\mathfrak{H}$, a construction that has been suggested in another form and connection in **(24)** for the finite-dimensional case, and which leads to difficulties of interpretation as pointed out there. On the other hand, the linearity of $\mathfrak{H}$ has ceased to play a role; the same construction can be made for any manifold $\mathfrak{H}$ (endowed with a suitable form Ω, determined in physics from the equations defining $\mathfrak{H}$). The approach therefore opens up a possible way of quantizing non-linear systems covariantly with respect to the group of all classical contact transformations.

References

1. W. Heisenberg and W. Pauli, *Quantum mechanics of wave fields*, Zeits. f. Physik, *56* (1929), 1–61.

2. N. Bohr and L. Rosenfeld, *Zur frage der Messbarkeit der Elektromagnetischen Feldgrossen*, Kgl. Danske Vidensk. Selsk., mat.-fys. Medd. XII, *8* (1953), 3–65.

3. V. Fock, *Konfigurationsraum und zweiten Quantelung*, Zeits. f. Physik, *75* (1952), 622–647.

4. K. O. Friedrichs, *Mathematical aspects of the quantum theory of fields, I-II*, Commun. Pure Appl. Math., *4* (1951), 161–224.

5. J. M. Cook, *The mathematics of second quantization*, Thesis, University of Chicago, 1951; in part in Trans. Amer. Math. Soc., *74* (1953), 222–245.

6. A. S. Wightman, *Quelques problèmes mathématiques de la théorie quantique relativiste*, in *Colloque* [Lille, 1957) *on the mathematical problems of quantum field theory* (Paris, 1959), 1–38.

7. G. Källén and A. S. Wightman, *The analytic properties of the vacuum expectation values of a product of three scalar fields*, K. Danske Vidensk. Selsk., mat-fys. Skr., *1* (1958).

8. J. A. Shohat and J. D. Tamarkin, *The problem of moments* (New York, 1943).

9A. G. Källén, Lectures at the Institute for Theoretical Physics, Lund, Spring, 1959.

9B. J. Schwinger, *Quantum electrodynamics*. I, *A covariant formalism*, Phys. Rev., *74* (1948), 1439–1461.

10. I. E. Segal, *Distributions in Hilbert space and canonical systems of operators*, Trans. Amer. Math. Soc., *88* (1958), 12–41.

11. I. E. Segal, *Direct formulation of causality requirements on the S-operator*, Phys. Rev., *109* (1958), 2191–2198.
12. ——— *Foundations of the theory of dynamical systems of infinitely many degrees of freedom, I*, Kgl. Danske Vidensk. Selsk., mat.-fys. Medd., *31* (1959), 1–38.
13. ——— *Postulates for general quantum mechanics*, Ann. Math., *48* (1947), 930–948.
14. F. J. Murray and J. von Neumann, *On rings of operators, IV*, Ann. Math., *44* (1943), 716–808.
15. I. Kaplansky, *A theorem on rings of operators*, Pac. J. Math., *1* (1951), 227–232.
16. I. E. Segal, *A non-commutative extension of abstract integration*, Ann. Math., *57* (1953), 401–457.
17. ——— *Tensor algebras over Hilbert spaces, I*, Trans. Amer. Math. Soc., *81* (1956), 106–134.
18. L. van Hove, *Sur certaines représentations unitaires d'un groupe infini de transformations*, Acad. Roy. Belgique, Cl. Sci. Mém. Coll in 8°, *26* (1951), 102 pp.
19. D. Shale, *On certain groups of operators on Hilbert space*, Doctoral Thesis, University of Chicago, 1959.
20. R. S. Phillips, *Integration in a convex linear topological space*, Trans. Amer. Math. Soc., *47* (1940), 114–145.
21. E. Wigner, *On the quantum correction for thermodynamic equilibrium*, Phys. Rev., *40* (1932), 749–759.
22. G. A. Baker, Jr., *Formulation of quantum mechanics based on the quasi-probability distribution induced in phase space*, Phys. Rev., *109* (1958), 2198–2206.
23. G. Hochschild, *Cohomology of restricted Lie algebras*, Amer. J. Math., *76* (1954), 555–580.
24. L. H. Thomas, *General relativity and particle dynamics*, Phys. Rev., *112* (1958), 2129–2134, and to appear.

RECIPROCAL CONVERGENCE CLASSES FOR FOURIER SERIES AND INTEGRALS

A. P. GUINAND

Introduction. The classical result of Plancherel for Fourier cosine transforms of functions $f(x)$ of the class $L^2(0, \infty)$ states that (see **(7)** for references)

$$g(x) = \underset{T\to\infty}{\text{l.i.m.}} \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \int_0^T f(t) \cos xt \, dt$$

converges in mean square to a function $g(x)$ which also belongs to $L^2(0, \infty)$, and furthermore

$$f(x) = \underset{T\to\infty}{\text{l.i.m.}} \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \int_0^T g(t) \cos xt \, dt.$$

Some years ago in a series of papers **(1; 2; 3)** on summation formulae I showed that a similar symmetrical theory for narrower classes of functions and ordinary convergence of the integrals can also be developed. The relevant results can be expressed as follows:

THEOREM 1. *If $f(x)$ is the integral of its derivative and $xf'(x)$ belongs to $L^2(0, \infty)$, then*

$$\lim_{x\to\infty} f(x) = l$$

exists, $f(x) - l$ belongs to $L^2(0, \infty)$, and

$$f(x) - l = o(x^{-\frac{1}{2}})$$

as x tends to $+0$ or to $+\infty$.

Definition 1. If $f(x)$ is the integral of its derivative and $x\,f'(x)$ belongs to $L^2(0, \infty)$, and if the limit to which $f(x)$ tends as x tends to infinity is zero, then we say that $f(x)$ belongs to the class $S_1^2(0, \infty)$.

THEOREM 2. *If $f(x)$ belongs to $S_1^2(0, \infty)$ then for $x > 0$*

$$g(x) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \int_0^{\to\infty} f(t) \cos xt \, dt \tag{1}$$

converges, $g(x)$ also belongs to $S_1^2(0, \infty)$, and

$$f(x) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \int_0^{\to\infty} g(t) \cos xt \, dt. \tag{2}$$

Here we use the notation

$$\int_a^{\to\infty} = \lim_{T\to\infty} \int_a^T.$$

Received September 20, 1959.

That is to say that the class $S_1^2(0, \infty)$ is a subclass of $L^2(0, \infty)$, and that it can be described as a self-reciprocal convergence class for Fourier cosine transformations.

This theory has recently been extended by Miller **(5)** to cover wider subclasses of $L^2(0, \infty)$ and more general transformations. A disadvantage of Theorem 2 is that, although the result is simple and easily applied, the proof of Theorem 2 is indirect and it uses results from the Plancherel theory.

In the first part of the present paper I show how to find narrower self-reciprocal convergence classes for Fourier cosine transforms, and I give a direct proof of the Fourier inversion formula without using the Plancherel theory for one such self-reciprocal convergence class.

In the second part of the paper I prove analogues of Theorems 1 and 2 for Fourier series. I define a class $S_1^2(0, 2\pi)$ of functions $f(x)$ of period 2π and a class $\sum_1^2 (-\infty, \infty)$ of sequences $\{c_n\}$ $(n = 0, \pm 1, \pm 2, \ldots,)$ which are reciprocal convergence classes for Fourier series in the sense that:

(i) if $f(x)$ belongs to $S_1^2(0, 2\pi)$ then it has a Fourier series

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{inx} \tag{3}$$

which converges for $x \not\equiv 0 \pmod{2\pi}$, and $\{c_n\}$ belongs to $\sum_1^2(-\infty, \infty)$;

(ii) if $\{c_n\}$ belongs to $\sum_1^2(-\infty, \infty)$ then $\sum_{n=-\infty}^{\infty} c_n e^{inx}$ converges for all $x \not\equiv 0 \pmod{2\pi}$ and defines a function $f(x)$ belonging to $S_1^2(0, 2\pi)$.

Part I: Fourier Integrals

1. Reciprocal classes and Mellin transforms. If $f(x)$ and $g(x)$ are Fourier cosine transforms connected by the equations (1) and (2), and $\mathfrak{F}(s)$ and $\mathfrak{G}(s)$ are their Mellin transforms then, formally **(7**, p. 213**)**,

$$\mathfrak{G}(s) = \mathfrak{K}(s)\, \mathfrak{F}(1 - s) \tag{4}$$

and

$$\mathfrak{F}(s) = \mathfrak{K}(s)\, \mathfrak{G}(1 - s),$$

where

$$\mathfrak{K}(s) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \Gamma(s) \cos \tfrac{1}{2} s\pi,$$

and consequently

$$|\mathfrak{K}(\tfrac{1}{2} + it)| = 1 \tag{5}$$

for all real t.

From the L^2 theory of Mellin transforms **(7**, p. 94**)** it follows that if $f(x)$ belongs to $L^2(0, \infty)$ then $\mathfrak{F}(s)$ belongs to $\mathfrak{L}^2(-\infty, \infty)$. Hence by (4) and (5) it follows that $\mathfrak{G}(s)$ also belongs to $\mathfrak{L}^2(-\infty, \infty)$ and consequently $g(x)$ belongs to $L^2(0, \infty)$, as required by the Plancherel theory.

A similar argument can be used to show that the class $S_1^2(0, \infty)$ is self-

reciprocal for Fourier cosine transformations. If $f(x)$ belongs to $S_1^2(0, \infty)$ then the Mellin transform of $x\,f'(x)$ exists and is

$$\begin{aligned}(6)\qquad & \underset{X\to\infty}{\text{l.i.m.}} \int_{1/X}^{X} xf'(x)x^{s-1}\,dx \\ &= \underset{X\to\infty}{\text{l.i.m.}} \left\{[x^s f(x)]_{1/X}^{X} - s\int_{1/X}^{X} f(x)x^{s-1}\,dx\right\} \\ &= -\,s\,\mathfrak{F}(s),\end{aligned}$$

since the integrated terms vanish for $R(s) = \frac{1}{2}$ by Theorem 1. Hence $s\,\mathfrak{F}(s)$ belongs to $\mathfrak{L}^2(-\infty, \infty)$ and it follows from (4) and (5) that $s\,\mathfrak{G}(s)$ also belongs to $\mathfrak{L}^2(-\infty, \infty)$. Then, reversing the above argument, it follows that $g(x)$ belongs to $S_1^2(0, \infty)$, as required by Theorem 2.

Now the same procedure can be used when $\mathfrak{F}(s)$, instead of being multiplied by $-s$ as in (6), is multiplied by some other suitable function of s. For example, put

$$\Phi(1 - s) = \mathfrak{F}(s)/\Gamma(s)$$

and assume that $\Phi(s)$ belongs to $\mathfrak{L}^2(-\infty, \infty)$. Then $\Phi(s)$ is the Mellin transform of a function $\phi(x)$ belonging to $L^2(0, \infty)$. Further $\Gamma(s)$ is the Mellin transform of e^{-x} and consequently the relationship

$$(7)\qquad \mathfrak{F}(s) = \Gamma(s)\Phi(1 - s)$$

corresponds to (**7**, p. 213)

$$(8)\qquad f(x) = \int_0^\infty e^{-xt}\phi(t)\,dt.$$

From (4) and (7) we have

$$\begin{aligned}\mathfrak{G}(s) &= \mathfrak{K}(s)\Gamma(1 - s)\Phi(s) \\ &= \left\{\mathfrak{K}(s)\frac{\Gamma(1 - s)}{\Gamma(s)}\right\}\{\Gamma(s)\Phi(s)\}.\end{aligned}$$

Hence, by (5), on $R(s) = \frac{1}{2}$

$$\left|\frac{\mathfrak{G}(s)}{\Gamma(s)}\right| = \left|\mathfrak{K}(s)\frac{\Gamma(1 - s)}{\Gamma(s)}\right| |\Phi(s)| = |\Phi(s)|$$

and so $\mathfrak{G}(s)/\Gamma(s)$ belongs to $\mathfrak{L}^2(-\infty, \infty)$.

Reversing the argument from (7) to (8) it follows that there is a $\psi(x)$ belonging to $L^2(0, \infty)$ for which

$$g(x) = \int_0^\infty e^{-xt}\psi(t)\,dt,$$

and we have the following result.

Theorem 3. *If $f(x)$ is the Laplace transform of a function of $L^2(0, \infty)$, and $g(x)$ is its Fourier cosine transform, then $g(x)$ is also the Laplace transform of a function of $L^2(0, \infty)$.*

Let us now make the following definition.

Definition 2. The function $f(x)$ is said to belong to the class $\Lambda^2\{h(x)\}$ if there exists a function $\phi(x)$ belonging to $L^2(0, \infty)$ such that

$$f(x) = \int_0^\infty h(xt)\phi(t)\,dt$$

for all $x > 0$.

Then Theorem 3 states that the class $\Lambda^2(e^{-x})$ is self-reciprocal for Fourier cosine transformations.

The same type of argument can be used to prove the following more general result.

Theorem 4. *If $h(x)$ belongs to $L^2(0, \infty)$ and has a Mellin transform $\mathfrak{H}(s)$ satisfying*

$$\left|\frac{\mathfrak{H}(\frac{1}{2} + it)}{\mathfrak{H}(\frac{1}{2} - it)}\right| = 1$$

for all real t then $\Lambda^2\{h(x)\}$ is a self-reciprocal class of functions with respect to any general transformation of the Fourier type.

For general transformations see (**7**, ch. VIII).

2. Symmetrical convergence theorems by direct methods. The arguments of § 1 do not prove that the Fourier integrals (1) and (2) converge, and they use the L^2 theory of Mellin transforms. If we consider the class of functions

$$\Lambda^2(e^{-\frac{1}{2}x^2})$$

we can derive a symmetrical convergence theorem for the Fourier cosine transformation by a direct method. The result is:

Theorem 5. *If $f(x)$ belongs to the class*

$$\Lambda^2(e^{-\frac{1}{2}x^2})$$

then

$$g(x) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \int_0^{\to\infty} f(t)\cos xt\,dt \tag{9}$$

converges for $x > 0$, $g(x)$ also belongs to

$$\Lambda^2(e^{-\frac{1}{2}x^2}),$$

and

$$f(x) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \int_0^{\to\infty} g(t) \cos xt \, dt \tag{10}$$

for $x > 0$. Further, if, for $x > 0$,

$$f(x) = \int_0^\infty e^{-\frac{1}{2}x^2 t^2} \phi(t) \, dt \tag{11}$$

and

$$g(x) = \int_0^\infty e^{-\frac{1}{2}x^2 t^2} \psi(t) \, dt$$

in accordance with Definition 2 then

$$\psi(x) = \frac{1}{x} \phi\left(\frac{1}{x}\right) \tag{12}$$

almost everywhere.

From Definition 2 there exists a function $\phi(x)$ of $L^2(0, \infty)$ satisfying (11). Hence

$$\begin{aligned} g(x) &= \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \int_0^\infty f(t) \cos xt \, dt \\ &= \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \int_0^\infty \cos xt \, dt \int_0^\infty \phi(u) e^{-\frac{1}{2}u^2 t^2} \, du \\ &= \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \int_0^\infty \phi(u) \, du \int_0^\infty e^{-\frac{1}{2}u^2 t^2} \cos xt \, dt, \end{aligned} \tag{13}$$

provided that this formal process can be justified. Now

$$\left(\frac{2}{\pi}\right)^{\frac{1}{2}} \int_0^\infty e^{-\frac{1}{2}u^2 t^2} \cos xt \, dt = \frac{1}{u} e^{-\frac{1}{2}x^2/u^2},$$

so (13) becomes

$$\begin{aligned} g(x) &= \int_0^\infty \phi(u) e^{-\frac{1}{2}x^2/u^2} \frac{du}{u} \\ &= \int_0^\infty \frac{1}{t} \phi\left(\frac{1}{t}\right) e^{-\frac{1}{2}x^2 t^2} \, dt \\ &= \int_0^\infty \psi(t) e^{-\frac{1}{2}x^2 t^2} \, dt \end{aligned} \tag{14}$$

by (12).

We can justify this process by the following three lemmas.

Lemma 1. *If $V_1 > V > 0$, $y > 0$, then*

$$\left| \int_V^{V_1} e^{-\frac{1}{2}v^2} \cos yv \, dv \right| < \frac{2}{y} e^{-\frac{1}{2}V^2}.$$

Proof.

$$\int_V^{V_1} e^{-\frac{1}{2}v^2} \cos yv \, dv = \left[\frac{\sin yv}{y} e^{-\frac{1}{2}v^2}\right]_V^{V_1} + \frac{1}{y}\int_V^{V_1} ve^{-\frac{1}{2}v^2} \sin yv \, dv.$$

Hence

$$\left|\int_V^{V_1} e^{-\frac{1}{2}v^2} \cos yv \, dv\right| < \frac{1}{y}(e^{-\frac{1}{2}V^2} + e^{-\frac{1}{2}V_1^2}) + \frac{1}{y}\int_V^{V_1} ve^{-\frac{1}{2}v^2} dv$$

$$= \frac{1}{y}(e^{-\frac{1}{2}V^2} + e^{-\frac{1}{2}V_1^2}) + \frac{1}{y}(e^{-\frac{1}{2}V^2} - e^{-\frac{1}{2}V_1^2})$$

$$= \frac{2}{y} e^{-\frac{1}{2}V^2}.$$

LEMMA 2. *If T, x, δ are positive real numbers and $\phi(x)$ belongs to $L^2(0, \infty)$ then*

$$\left|\int_T^\infty \cos xt \, dt \int_0^\delta e^{-\frac{1}{2}u^2t^2} \phi(u) du\right| < \frac{2^{\frac{1}{2}}\pi^{\frac{1}{4}}}{xT^{\frac{1}{2}}} \left\{\int_0^\delta |\phi(u)|^2 du\right\}^{\frac{1}{2}}. \tag{15}$$

Proof. Consider

$$\left|\int_T^{T_1} \cos xt \, dt \int_0^\delta e^{-\frac{1}{2}u^2t^2} \phi(u) \, du\right| \tag{16}$$

where $T_1 > T > 0$. Since $\phi(x)$ belongs to $L^2(0, \infty)$ it follows that $\phi(x)$ belongs to $L(0, \delta)$ for any finite δ, and that (16) converges absolutely. Hence (16) is equal to

$$\left|\int_0^\delta \phi(u) \, du \int_T^{T_1} e^{-\frac{1}{2}u^2t^2} \cos ut \, dt\right| \tag{17}$$

$$= \left|\int_0^\delta \phi(u) \frac{du}{u} \int_{Tu}^{T_1u} e^{-\frac{1}{2}v^2} \cos \frac{xv}{u} dv\right|$$

$$< \frac{2}{x}\int_0^\delta |\phi(u)| e^{-\frac{1}{2}T^2u^2} du$$

by Lemma 1. By Schwarz's inequality (17) is less than or equal to

$$\frac{2}{x}\left\{\int_0^\delta |\phi(u)|^2 du\right\}^{\frac{1}{2}} \left\{\int_0^\delta e^{-T^2u^2} du\right\}^{\frac{1}{2}} \tag{18}$$

$$< \frac{2}{x}\left\{\int_0^\delta |\phi(u)|^2 du\right\}^{\frac{1}{2}} \left\{\int_0^\infty e^{-T^2u^2} du\right\}^{\frac{1}{2}}$$

$$= \frac{2^{\frac{1}{2}}\pi^{\frac{1}{4}}}{xT^{\frac{1}{2}}}\left\{\int_0^\delta |\phi(u)|^2 du\right\}^{\frac{1}{2}}.$$

Making T and T_1 tend to infinity it follows that the double integral in (15) converges, and (15) follows from (18) if we keep T fixed and make T_1 tend to infinity.

LEMMA 3. *If x is real and positive, and $\phi(x)$ belongs to $L^2(0, \infty)$ then*

$$\int_0^{\to\infty} \cos xt\, dt \int_0^\infty \phi(u) e^{-\frac{1}{2}u^2t^2} du \tag{19}$$

converges and is equal to

$$\int_0^\infty \phi(u)\, du \int_0^\infty e^{-\frac{1}{2}u^2t^2} \cos xt\, dt. \tag{20}$$

Proof. The inversion of order of integration

$$\int_\delta^\infty \phi(u)\, du \int_0^\infty e^{-\frac{1}{2}u^2t^2} \cos xt\, dt = \int_0^\infty \cos xt\, dt \int_\delta^\infty \phi(u)\, e^{-\frac{1}{2}u^2t^2}\, du$$

is justified by absolute convergence since

$$\begin{aligned}
\int_\delta^\infty |\phi(u)| du \int_0^\infty |e^{-\frac{1}{2}u^2t^2} \cos xt| dt &< \int_\delta^\infty |\phi(u)| du \int_0^\infty e^{-\frac{1}{2}u^2t^2}\, dt \\
&= (2\pi)^{\frac{1}{2}} \int_\delta^\infty |\phi(u)|\, \frac{du}{u} \\
&< (2\pi)^{\frac{1}{2}} \left\{\int_\delta^\infty |\phi(u)|^2\, du\right\}^{\frac{1}{2}} \left\{\int_\delta^\infty \frac{du}{u^2}\right\}^{\frac{1}{2}} \\
&= \left(\frac{2\pi}{\delta}\right)^{\frac{1}{2}} \left\{\int_\delta^\infty |\phi(u)|^2\, du\right\}^{\frac{1}{2}}.
\end{aligned}$$

Also

$$\begin{aligned}
\int_0^\delta |\phi(u)| e^{-\frac{1}{2}x^2/u^2} \frac{du}{u} &< \left\{\int_0^\delta |\phi(u)|^2\, du\right\}^{\frac{1}{2}} \left\{\int_0^\delta e^{-x^2/u^2} \frac{du}{u^2}\right\}^{\frac{1}{2}} \\
&= \left\{\int_0^\delta |\phi(u)|^2\, du\right\}^{\frac{1}{2}} \left\{\int_{1/\delta}^\infty e^{-x^2v^2}\, dv\right\}^{\frac{1}{2}} \\
&< \left\{\int_0^\delta |\phi(u)|^2\, du\right\}^{\frac{1}{2}} \left\{\int_0^\infty e^{-x^2v^2}\, dv\right\}^{\frac{1}{2}} \\
&= 2^{-\frac{1}{2}}\pi^{\frac{1}{4}}x^{-\frac{1}{2}} \left\{\int_0^\delta |\phi(u)|^2\, du\right\}^{\frac{1}{2}}.
\end{aligned} \tag{21}$$

Hence

$$\begin{aligned}
&\int_0^{\to\infty} \cos xt\, dt \int_0^\infty \phi(u)\, e^{-\frac{1}{2}u^2t^2}\, du - \int_0^\infty \phi(u)\, du \int_0^\infty e^{-\frac{1}{2}u^2t^2} \cos xt\, dt \\
&\quad = \int_0^{\to\infty} \cos xt\, dt \int_0^\infty \phi(u)\, e^{-\frac{1}{2}u^2t^2}\, du - \int_0^\delta \phi(u)\, du \int_0^\infty e^{-\frac{1}{2}u^2t^2} \cos xt\, dt \\
&\qquad\qquad - \int_\delta^\infty \phi(u)\, du \int_0^\infty e^{-\frac{1}{2}u^2t^2} \cos xt\, dt \\
&\quad = \int_0^{\to\infty} \cos xt\, dt \int_0^\delta \phi(u)\, e^{-\frac{1}{2}u^2t^2}\, du - \int_0^\delta \phi(u)\, du \int_0^\infty e^{-\frac{1}{2}u^2t^2} \cos xt\, dt \\
&\quad = \int_0^T \cos xt\, dt \int_0^\delta \phi(u)\, e^{-\frac{1}{2}u^2t^2}\, du + \int_T^{\to\infty} \cos xt\, dt \int_0^\delta \phi(u)\, e^{-\frac{1}{2}u^2t^2}\, du \\
&\qquad\qquad - \left(\frac{\pi}{2}\right)^{\frac{1}{2}} \int_0^\delta \phi(u)\, e^{-\frac{1}{2}x^2/u^2} \frac{du}{u}.
\end{aligned} \tag{22}$$

Now

$$
\begin{aligned}
(23) \qquad \left| \int_0^T \cos xt\, dt \int_0^\delta \phi(u)\, e^{-\frac{1}{2}u^2t^2}\, du \right|
&< \int_0^T dt \int_0^\delta |\phi(u)|\, du \\
&< T \left\{ \int_0^\delta |\phi(u)|^2\, du \right\}^{\frac{1}{2}} \left\{ \int_0^\delta du \right\}^{\frac{1}{2}} \\
&= T\delta^{\frac{1}{2}} \left\{ \int_0^\delta |\phi(u)|^2\, du \right\}^{\frac{1}{2}}.
\end{aligned}
$$

Hence, by Lemma 2, (21), (22), and (23)

$$
\begin{aligned}
&\left| \int_0^{\to\infty} \cos xt\, dt \int_0^\infty \phi(u)\, e^{-\frac{1}{2}u^2t^2}\, du - \int_0^\infty \phi(u)\, du \int_0^\infty e^{-\frac{1}{2}u^2t^2} \cos xt\, dt \right| \\
&\qquad < \left\{ T\delta^{\frac{1}{2}} + \frac{2^{\frac{1}{2}}\pi^{\frac{1}{2}}}{xT^{\frac{1}{2}}} + 2^{-\frac{1}{2}}\pi^{\frac{1}{2}}x^{-\frac{1}{2}} \right\} \left\{ \int_0^\delta |\phi(u)|^2\, du \right\}^{\frac{1}{2}}.
\end{aligned}
$$

This can be made arbitrarily small by choosing T first and then making δ sufficiently small, and Lemma 3 follows.

Proof of Theorem 5. Lemma 3 justifies the result (14). Further $\psi(x)$ belongs to $L^2(0, \infty)$ since

$$
\begin{aligned}
\int_0^\infty |\psi(x)|^2\, dx &= \int_0^\infty \frac{1}{x^2} \left| \phi\left(\frac{1}{x}\right) \right|^2 dx \\
&= \int_0^\infty |\phi(u)|^2\, du.
\end{aligned}
$$

Hence $g(x)$, defined by (9), belongs to $\Lambda^2(e^{-\frac{1}{2}x^2})$. Then repeating the preceding argument

$$
\begin{aligned}
\left(\frac{2}{\pi}\right)^{\frac{1}{2}} \int_0^{\to\infty} g(t) \cos xt\, dt &= \int_0^\infty \frac{1}{t}\, \psi\left(\frac{1}{t}\right) e^{-\frac{1}{2}x^2t^2}\, dt \\
&= \int_0^\infty \phi(t)\, e^{-\frac{1}{2}x^2t^2}\, dt \\
&= f(x)
\end{aligned}
$$

since (12) implies that for almost all x

$$
\phi(x) = \frac{1}{x}\, \psi\left(\frac{1}{x}\right).
$$

This completes the proof of Theorem 5.

Part II: Fourier Series

3. The class of functions $S_1^2(0, 2\pi)$.

Definition 3. If $f(x)$ is a periodic function of period 2π, is the integral of its

derivative, and $(\sin \frac{1}{2} x) f'(x)$ belongs to $L^2(0, 2\pi)$, then we say that $f(x)$ belongs to the class $S_1^2(0, 2\pi)$.

Definition 4. If $f(x)$ is a periodic function of period 2π, and is such that there exists a function $\phi(x)$ of $L^2(0, 2\pi)$ for which

$$f(x) = \operatorname{cosec} \tfrac{1}{2} x \int_0^x \phi(t)\, dt \tag{24}$$

and

$$\int_0^{2\pi} \phi(t)\, dt = 0, \tag{25}$$

then we say that $f(x)$ belongs to the class $S_1^2[0, 2\pi]$.

These definitions give two ways of characterizing the same class of functions. Properties of this class of functions are given by the following theorem.

THEOREM 6. *The classes $S_1^2(0, 2\pi)$ and $S_1^2[0, 2\pi]$ are identical, and all functions $f(x)$ of either class belong to $L^2(0, 2\pi)$. Also $x^{\frac{1}{2}} f(x)$ and $x^{\frac{1}{2}} f(2\pi - x)$ both tend to zero as $x \to +0$.*

This result is analogous to results given by **(4)**.
To prove the result we use Lemmas 4, 5, and 6.

LEMMA 4. *If $f(x)$ belongs to $S_1^2[0, 2\pi]$ then $x^{\frac{1}{2}} f(x)$ and $x^{\frac{1}{2}} f(2\pi - x)$ both tend to zero as $x \to +0$, and $f(x)$ belongs to $L^2(0, 2\pi)$.*

Proof. As $x \to +0$, by (24) and Schwarz's inequality

$$\begin{aligned} |f(x)|^2 &< \operatorname{cosec}^2 \tfrac{1}{2} x \left\{\int_0^x |\phi(t)|^2\, dt\right\}\left\{\int_0^x dt\right\} \\ &= o(x^{-1}). \end{aligned}$$

Hence $x^{\frac{1}{2}} f(x) \to 0$ as $x \to +0$. Further

$$\begin{aligned} f(2\pi - x) &= \operatorname{cosec} \tfrac{1}{2} x \int_0^{2\pi - x} \phi(t)\, dt \\ &= -\operatorname{cosec} \tfrac{1}{2} x \int_{2\pi - x}^{2\pi} \phi(t)\, dt \end{aligned}$$

by (25), and a similar argument shows that $x^{\frac{1}{2}} f(2\pi - x) \to 0$ as $x \to +0$. Now let $0 < a < b < 2\pi$, put

$$\phi_1(x) = \int_0^x \phi(t)\, dt,$$

and suppose that $f(x)$ is real. Then

$$\int_a^b \{f(x)\}^2\, dx = \int_a^b \operatorname{cosec}^2 \tfrac{1}{2} x\, \{\phi_1(x)\}^2\, dx. \tag{26}$$

Integrating by parts (26) becomes

$$[-2\cot \tfrac{1}{2}x\{\phi_1(x)\}^2]_a^b + 4\int_a^b \cot \tfrac{1}{2}x\, \phi(x)\phi_1(x)\, dx.$$

As $a \to +0$ and $b \to 2\pi - 0$, this is

$$o(1) + 4\int_a^b \cos \tfrac{1}{2}x\, \phi(x) f(x)\, dx$$

$$< o(1) + 4\left\{\int_a^b \cos^2 \tfrac{1}{2}x\, |\phi(x)|^2\, dx\right\}^{\frac{1}{2}} \left\{\int_a^b |f(x)|^2\, dx\right\}^{\frac{1}{2}}$$

$$< o(1) + 4\left\{\int_0^{2\pi} |\phi(x)|^2\, dx\right\}^{\frac{1}{2}} \left\{\int_a^b |f(x)|^2\, dx\right\}^{\frac{1}{2}}.$$

Dividing by

$$\left\{\int_a^b |f(x)|^2\, dx\right\}^{\frac{1}{2}}$$

and taking the limit as $a \to +0$, $b \to 2\pi - 0$, we have

$$\left\{\int_0^{2\pi} |f(x)|^2\, dx\right\}^{\frac{1}{2}} < 4\left\{\int_0^{2\pi} |\phi(x)|^2\, dx\right\}^{\frac{1}{2}},$$

and hence $f(x)$ belongs to $L^2(0, 2\pi)$, as required. If $f(x)$ is complex the result follows by splitting into real and imaginary parts.

LEMMA 5. *If $f(x)$ belongs to $S_1^2(0, 2\pi)$ then $x^{\frac{1}{2}} f(x)$ and $x^{\frac{1}{2}} f(2\pi - x)$ both tend to zero as $x \to +0$, and $f(x)$ belongs to $L^2(0, 2\pi)$.*

Proof. By Definition 3 there exists a function $\psi(x) = \sin \tfrac{1}{2}x\, f'(x)$, belonging to $L^2(0, 2\pi)$, such that

$$f(\pi) - f(x) = \int_x^\pi \operatorname{cosec} \tfrac{1}{2}t\, \psi(t)\, dt. \tag{27}$$

Suppose that $f(\pi) = 0$ and consider the behaviour of $f(x)$ as $x \to +0$. Choose $0 < \delta < \pi$ so that

$$\int_0^\delta |\psi(t)|^2\, dt < \epsilon.$$

Then for $0 < x < \delta$

$$|f(x)| < \int_\delta^\pi \operatorname{cosec} \tfrac{1}{2}t |\psi(t)| dt + \int_x^\delta \operatorname{cosec} \tfrac{1}{2}t |\psi(t)| dt$$

$$< \int_\delta^\pi \operatorname{cosec} \tfrac{1}{2}t |\psi(t)| dt + \left\{\int_x^\delta |\psi(t)|^2\, dt\right\}^{\frac{1}{2}} \left\{\int_x^\delta \operatorname{cosec}^2 \tfrac{1}{2}t\, dt\right\}^{\frac{1}{2}}$$

$$< \operatorname{cosec} \tfrac{1}{2}\delta \int_0^\pi |\psi(t)| dt + \epsilon^{\frac{1}{2}}(2\cot \tfrac{1}{2}x - 2\cot \tfrac{1}{2}\delta)^{\frac{1}{2}}.$$

Hence as $x \to +0$

$$\begin{aligned} x^{\frac{1}{2}}f(x) &= O(x^{\frac{1}{2}}) + O(\epsilon^{\frac{1}{2}})(x\cot\tfrac{1}{2}x - x\cot\tfrac{1}{2}\delta)^{\frac{1}{2}} \\ &= O(x^{\frac{1}{2}}) + O(\epsilon^{\frac{1}{2}}) \\ &= o(1) \end{aligned}$$

since $x \cot \frac{1}{2} x \to 2$ as $x \to 0$.

A similar argument also shows that $x^{\frac{1}{2}} f(2\pi - x) \to 0$ as $x \to +0$.

Now suppose that $0 < a < \pi$, and that $f(x)$ is real, and put

$$f_1(x) = \int_0^x f(t)\, dt.$$

Hence $f_1(x) = o(x^{\frac{1}{2}})$ as $x \to +0$. Also

$$\begin{aligned} |f(x)| &= \left| \int_x^\pi \operatorname{cosec} \tfrac{1}{2} t\, \psi(t)\, dt \right| \\ &< \left\{ \int_x^\pi |\psi(t)|^2\, dt \right\}^{\frac{1}{2}} \left\{ \int_x^\pi \operatorname{cosec}^2 \tfrac{1}{2} t\, dt \right\}^{\frac{1}{2}} \\ &< \left\{ \int_0^\pi |\psi(t)|^2\, dt \right\}^{\frac{1}{2}} \left\{ 2 \cot \tfrac{1}{2} x \right\}^{\frac{1}{2}}, \end{aligned}$$

whence $f_1(x)$ is bounded for the whole interval $(0, \pi)$.

Now let $0 < a < \pi$ and consider

$$\begin{aligned} \int_a^\pi |f(x)|^2\, dx &= \int_a^\pi f(x)\, dx \int_x^\pi \operatorname{cosec} \tfrac{1}{2} t\, \psi(t)\, dt \\ &= \Big[f_1(x)\, f(x) \Big]_a^\pi + \int_a^\pi f_1(x) \operatorname{cosec} \tfrac{1}{2} x\, \psi(x)\, dx. \end{aligned}$$

As $x \to \pi - 0$, $f_1(x)$ is bounded and $f(x) \to f(\pi) = 0$, and as $x \to +0$

$$f_1(x) f(x) = o(x^{\frac{1}{2}}) o(x^{-\frac{1}{2}}) = o(1).$$

Hence

$$\begin{aligned} (28) \quad \int_a^\pi \{f(x)\}^2 dx &= o(1) + \int_a^\pi f_1(x) \operatorname{cosec} \tfrac{1}{2} x\, \psi(x)\, dx \\ &\leqslant o(1) + \left\{ \int_a^\pi \left| f_1(x) \operatorname{cosec} \tfrac{1}{2} x \right|^2 dx \right\}^{\frac{1}{2}} \left\{ \int_a^\pi |\psi(x)|^2\, dx \right\}^{\frac{1}{2}}. \end{aligned}$$

Now

$$\begin{aligned} f_1(x) &= \int_0^x f(t)\, dt \\ &= \int_0^x dt \int_t^\pi \operatorname{cosec} \tfrac{1}{2} u\, \psi(u)\, du \\ &= \int_0^x \operatorname{cosec} \tfrac{1}{2} u\, \psi(u)\, du \int_0^u dt + \int_x^\pi \operatorname{cosec} \tfrac{1}{2} u\, \psi(u)\, du \int_0^x dt \\ &= \int_0^x u \operatorname{cosec} \tfrac{1}{2} u\, \psi(u)\, du - x f(x) \end{aligned}$$

by (27). Hence

(29) $$f_1(x) \operatorname{cosec} \tfrac{1}{2} x = \operatorname{cosec} \tfrac{1}{2} x \int_0^x u \operatorname{cosec} \tfrac{1}{2} u\, \psi(u)\, du - x \operatorname{cosec} \tfrac{1}{2} x\, f(x).$$

Now

$$2 < x \operatorname{cosec} \tfrac{1}{2}x < \pi$$

for $0 < x < \pi$. Hence $x \operatorname{cosec} \frac{1}{2}x\, \psi(x)$ belongs to $L^2(0, \pi)$, and by Lemma 4

(30) $$\operatorname{cosec} \tfrac{1}{2} x \int_0^x u \operatorname{cosec} \tfrac{1}{2} u\, \psi(u)\, du$$

belongs to $L^2(0, \pi)$.

Substituting (29) in (28) and using Minkowski's inequality we have

$$\int_a^\pi |f(x)|^2\, dx < o(1) + \left[\left\{\int_a^\pi \left|\operatorname{cosec} \tfrac{1}{2} x \int_0^x u \operatorname{cosec} \tfrac{1}{2} u \psi(u) du\right|^2 dx\right\}^{\frac{1}{2}}\right.$$
$$\left. + \left\{\int_a^\pi \pi^2 |f(x)|^2\, dx\right\}^{\frac{1}{2}}\right] \times \left\{\int_a^\pi |\psi(x)|^2\, dx\right\}^{\frac{1}{2}}$$
$$< o(1) + \left[\left\{\int_0^\pi \left|\operatorname{cosec} \tfrac{1}{2} x \int_0^x u \operatorname{cosec} \tfrac{1}{2} u\, \psi(u)\, du\right|^2 dx\right\}^{\frac{1}{2}}\right.$$
$$\left. + \pi\left\{\int_a^\pi |f(x)|^2\, dx\right\}^{\frac{1}{2}}\right] \times \left\{\int_0^\pi |\psi(x)|^2\, dx\right\}^{\frac{1}{2}}$$

since we know that (30) and $\psi(x)$ both belong to $L^2(0, \pi)$. That is,

(31) $$\int_a^\pi |f(x)|^2\, dx < A + B\left\{\int_a^\pi |f(x)|^2\, dx\right\}^{\frac{1}{2}}$$

where A and B are constants independent of a. Now unless $f(x)$ vanishes almost everywhere in $(0, \pi)$ we can find an a_1 and a k such that $0 < a_1 < \pi$ and

$$\left\{\int_{a_1}^\pi |f(x)|^2\, dx\right\}^{\frac{1}{2}} > k > 0.$$

From (31)

$$\left\{\int_a^\pi |f(x)|^2\, dx\right\}^{\frac{1}{2}} < A\left\{\int_a^\pi |f(x)|^2\, dx\right\}^{-\frac{1}{2}} + B$$
$$< \frac{A}{k} + B$$

for $a < a_1$. Hence

$$\left\{\int_0^\pi |f(x)|^2\, dx\right\}^{\frac{1}{2}} < \frac{A}{k} + B$$

and $f(x)$ belongs to $L^2(0, \pi)$. Combining the above with a similar argument for the interval $(\pi, 2\pi)$ we find that $f(x)$ belongs to $L^2(0, 2\pi)$.

If $f(\pi)$ is not zero the above argument shows that $f(x) - f(\pi)$ belongs to $L^2(0, 2\pi)$, so $f(x)$ belongs to $L^2(0, 2\pi)$.

Lastly, if $f(x)$ is complex the result of Lemma 5 follows by splitting into real and imaginary parts.

LEMMA 6. *The classes $S_1^2(0, 2\pi)$ and $S_1^2[0, 2\pi]$ are identical.*

Proof. With $\phi(x)$ and $\psi(x)$ as in Lemmas 4 and 5 we have

$$f(x) = \operatorname{cosec} \tfrac{1}{2} x \int_0^x \phi(t)\, dt = f(\pi) - \int_x^\pi \operatorname{cosec} \tfrac{1}{2} t\, \psi(t)\, dt. \tag{32}$$

By differentiation

$$\phi(x) = \sin \tfrac{1}{2}x\, f'(x) + \tfrac{1}{2} \cos \tfrac{1}{2}x\, f(x)$$

and

$$\psi(x) = \sin \tfrac{1}{2}x\, f'(x)$$

almost everywhere in $(0, 2\pi)$. Hence

$$\phi(x) = \psi(x) + \tfrac{1}{2} \cos \tfrac{1}{2}x\, f(x) \tag{33}$$

almost everywhere in $(0, 2\pi)$.

Now if $f(x)$ belongs to $S_1^2\,[0, 2\pi]$ this means that $\phi(x)$ belongs to $L^2(0, 2\pi)$, and, by Lemma 4, so does $f(x)$. Hence from (33) $\psi(x)$ also belongs to $L^2(0, 2\pi)$; that is, $f(x)$ belongs to $S_1^2(0, 2\pi)$.

Conversely if $f(x)$ belongs to $S_1^2(0, 2\pi)$ then by Lemma 5 it also belongs to $L^2(0, 2\pi)$ and $\psi(x)$ belongs to $L^2(0, 2\pi)$. Hence from (33) $\phi(x)$ also belongs to $L^2(0, 2\pi)$. Also by (32)

$$\int_0^x \phi(t)\, dt = \sin \tfrac{1}{2}\, x\, f(x).$$

By Lemma 5 $x^{\frac{1}{2}} f(2\pi - x) \to 0$ as $x \to +0$. Hence

$$\begin{aligned}\int_0^{2\pi} \phi(t)\, dt &= \lim_{x \to +0} \{\sin \tfrac{1}{2}(2\pi - x)\, f(2\pi - x)\} \\ &= 0.\end{aligned}$$

That is, $f(x)$ belongs to $S_1^2[0, 2\pi]$, and this completes the proof of Lemma 6.
Combining Lemmas 4, 5, and 6 we have Theorem 6.

We also require the following result to connect $S_1^2(0, 2\pi)$ with Fourier Series in § 5.

THEOREM 7. *The class $S_1^2(0, 2\pi)$ is identical with the class of functions $f(x)$ of period 2π which can be expressed in the form*

$$f(x) = \frac{1}{1 - e^{-ix}} \int_0^x \chi(t)\, dt, \tag{34}$$

where $\chi(x)$ belongs to $L^2(0, 2\pi)$ and

$$\int_0^{2\pi} \chi(t)\, dt = 0. \tag{35}$$

Proof. By (24), (25), (34), and (35) we require that

$$\frac{1}{1-e^{-ix}}\int_0^x \chi(t)\,dt = \operatorname{cosec}\tfrac{1}{2}x\int_0^x \phi(t)\,dt \tag{36}$$

where

$$\int_0^{2\pi} \phi(t)\,dt = \int_0^{2\pi} \chi(t)\,dt = 0$$

and $\phi(x)$ and $\chi(x)$ belong to $L^2(0, 2\pi)$. Now (36) gives

$$\int_0^x \chi(t)\,dt = 2i\,e^{-\frac{1}{2}ix}\int_0^x \phi(t)\,dt \tag{37}$$

and hence

$$\chi(x) = 2i\,e^{-\frac{1}{2}ix}\,\phi(x) + e^{\frac{1}{2}ix}\int_0^x \phi(t)\,dt \tag{38}$$

almost everywhere in $(0, 2\pi)$. If $\phi(x)$ belongs to $L^2(0, 2\pi)$, so does

$$\int_0^x \phi(t)\,dt,$$

and hence from (38), $\chi(x)$ belongs to $L^2(0, 2\pi)$. A similar argument shows that $\phi(x)$ belongs to $L^2(0, 2\pi)$ if $\chi(x)$ does.

Finally if we put $x = 2\pi$ in (37) we have

$$\int_0^{2\pi} \chi(t)\,dt = -2i\int_0^{2\pi} \phi(t)\,dt.$$

Hence the vanishing of either of these integrals implies the vanishing of the other.

4. The class of sequences $\sum_1^2(-\infty, \infty)$

THEOREM 8. *If* $\{c_n\}$, $(n = 0, 1, 2, \ldots,)$ *is a sequence of complex numbers such that the series*

$$\sum_{n=1}^{\infty} n^2|c_n - c_{n+1}|^2$$

is convergent, then

(i) c_n *tends to a finite limit* l *as* $n \to \infty$, *and*

$$c_n - l = o(n^{-\frac{1}{2}}), \tag{39}$$

(ii) *the series*

$$\sum_{n=0}^{\infty} |c_n|^2$$

converges.

Proof of (i). If $m > n > 1$ then

$$(40) \qquad |c_n - c_m| = \left| \sum_{r=n}^{m-1} (c_r - c_{r+1}) \right|$$
$$< \left\{ \sum_{r=n}^{m-1} r^2 |c_r - c_{r+1}|^2 \right\}^{\frac{1}{2}} \left\{ \sum_{r=n}^{m-1} r^{-2} \right\}^{\frac{1}{2}}$$
$$= o(n^{-\frac{1}{2}}).$$

Hence, by the principle of convergence, c_n tends to a finite limit l as $n \to \infty$, and making $m \to \infty$ in (40) we have (39).

LEMMA 7. *If* $\{a_n\}$, $(n = 1, 2, 3, \ldots,)$ *is any sequence of complex numbers and* N *a positive integer then*

$$6 \sum_{n=1}^{N-1} n^2 |a_n - a_{n+1}|^2 + 2N|a_N|^2 \geq \sum_{n=1}^{N} |a_n|^2.$$

Proof of Lemma 7. We have

$$\begin{aligned} 6n^2|a_n - a_{n+1}|^2 + (2n+1)|a_{n+1}|^2 &= (2n^2 - 2n)|a_n - a_{n+1}|^2 + \{2n|a_n - a_{n+1}| - |a_{n+1}|\}^2 \\ &\qquad + 2n\{|a_n - a_{n+1}| + |a_{n+1}|\}^2 \\ &\geq 2n\{|a_n - a_{n+1}| + |a_{n+1}|\}^2 \\ &\geq 2n|a_n|^2 \end{aligned}$$

since $2n^2 - 2n \geq 0$ for all integers n and

$$|a_n - a_{n+1}| + |a_{n+1}| \geq |a_n|.$$

Hence

$$6n^2|a_n - a_{n+1}|^2 + 2(n+1)|a_{n+1}|^2 - 2n|a_n|^2 \geq |a_{n+1}|^2,$$

and the lemma follows on summing over $n = 0, 1, 2, \ldots, N-1$.

Proof of (ii). If we put $a_n = c_n - l$ then by (39) $N|a_N|^2$ tends to zero as $N \to \infty$. Hence

$$6 \sum_{n=1}^{\infty} n^2 |c_n - c_{n+1}|^2 \geq \sum_{n=1}^{\infty} |c_n - l|^2$$

and the latter series converges.

Definition 5. If $\{c_n\}$, $(n = 0, \pm 1, \pm 2, \ldots,)$ is a sequence of complex numbers such that

$$\sum_{n=-\infty}^{\infty} n^2 |c_n - c_{n+1}|^2$$

converges, and if the limits to which c_n tends as $n \to \pm\infty$ are both zero, then we say that the sequence $\{c_n\}$ belongs to the class $\sum_1{}^2(-\infty, \infty)$.

5. The convergence of Fourier series for $S_1^2(0, 2\pi)$.

THEOREM 9. *If* $\{c_n\}$ *belongs to the class* $\sum_1^2(-\infty, \infty)$ *then the series*

$$\sum_{n=-\infty}^{\infty} c_n e^{inx} \tag{41}$$

converges for all x not congruent to zero modulo 2π, *and its sum defines a function* $f(x)$, *belonging to* $S_1^2(0, 2\pi)$, *of which* (41) *is the Fourier series.*

Proof. Consider the series

$$\sum_{n=1}^{\infty} c_n e^{inx}, \tag{42}$$

and put

$$n(c_n - c_{n+1}) = \chi_n.$$

Then

$$\begin{aligned} c_n &= (c_n - c_{n+1}) + (c_{n+1} - c_{n+2}) + \dots \\ &= \sum_{r=n}^{\infty} \frac{\chi_r}{r}. \end{aligned}$$

Hence

$$\begin{aligned} \sum_{n=1}^{N} c_n e^{inx} &= \sum_{n=1}^{N} e^{inx} \sum_{r=n}^{\infty} \frac{\chi_r}{r} \\ &= \sum_{r=1}^{N} \frac{\chi_r}{r} \sum_{n=1}^{r} e^{inx} + \sum_{r=N+1}^{\infty} \frac{\chi_r}{r} \sum_{n=1}^{N} e^{inx} \\ &= \sum_{r=1}^{N} \frac{\chi_r}{r} \left\{ \frac{e^{i(r+1)x} - 1}{e^{ix} - 1} \right\} + \sum_{r=N+1}^{\infty} \frac{\chi_r}{r} \left\{ \frac{e^{i(N+1)x} - 1}{e^{ix} - 1} \right\}. \end{aligned} \tag{43}$$

Now

$$\begin{aligned} \sum_{r=N+1}^{\infty} \left| \frac{\chi_r}{r} \right| &< \left\{ \sum_{r=N+1}^{\infty} |\chi_r|^2 \right\}^{\frac{1}{2}} \left\{ \sum_{r=N+1}^{\infty} r^{-2} \right\}^{\frac{1}{2}} \\ &= o(N^{-\frac{1}{2}}) \end{aligned}$$

since the series

$$\sum_{r=1}^{\infty} |\chi_r|^2 \tag{44}$$

converges by hypothesis. Hence by (43), if $x \not\equiv 0 \bmod 2\pi$

$$\sum_{n=1}^{N} c_n e^{inx} = \frac{1}{e^{ix} - 1} \sum_{r=1}^{N} \frac{\chi_r}{r} \{e^{i(r+1)x} - 1\} + o(N^{-\frac{1}{2}}),$$

and the series on the right is absolutely convergent. Hence the series (42) converges, and

$$\sum_{n=1}^{\infty} c_n e^{inx} = \frac{1}{e^{ix} - 1} \sum_{r=1}^{\infty} \frac{\chi_r}{r} \{e^{i(r+1)x} - 1\}. \tag{45}$$

Since the series (44) converges the series

$$\sum_{r=1}^{\infty} \chi_r e^{irx}$$

converges in mean square to a function $\chi(x)$ belonging to $L^2(0, 2\pi)$, and is the Fourier series of this function. Hence, by the Fourier series integration theorem (**6**, p. 419),

$$\int_0^x \chi(t)dt = -i \sum_{r=1}^{\infty} \frac{\chi_r}{r} (e^{irx} - 1),$$

and in particular

$$\int_0^{2\pi} \chi(t)dt = 0.$$

Hence by (45)

$$\sum_{n=1}^{\infty} c_n e^{inx} = \frac{i}{1 - e^{-ix}} \int_0^x \chi(t)\, dt + \sum_{r=1}^{\infty} \frac{\chi_r}{r}. \tag{46}$$

By Theorem 7 it follows that (46) is a function of the class $S_1^2(0, 2\pi)$. A similar argument for negative n shows that the whole series (41) converges for $x \not\equiv 0 \bmod 2\pi$, and that its sum is a function of the class $S_1^2(0, 2\pi)$.

Since $S_1^2(0, 2\pi)$ is a subclass of $L^2(0, 2\pi)$ the series (41) must be the Fourier series of its sum.

THEOREM 10. *If $f(x)$ belongs to the class $S_1^2(0, 2\pi)$ then it has a Fourier series*

$$\sum_{n=-\infty}^{\infty} c_n e^{inx}$$

which converges to $f(x)$ for all x not congruent to zero modulo 2π, and the sequence $\{c_n\}$ belongs to the class $\sum_1^2(-\infty, \infty)$.

Proof. By Theorem 6 the function $f(x)$ belongs to $L^2(0, 2\pi)$. Hence it has a Fourier series

$$f(x) \sim \sum_{n=-\infty}^{\infty} c_n e^{inx} \tag{47}$$

for which c_n tends to zero as $n \to \pm\infty$.

By Theorem 7 there exists a function $\chi(x)$ of $L^2(0, 2\pi)$ such that

$$f(x) = \frac{i}{1 - e^{-ix}} \int_0^x \chi(t)\, dt \tag{48}$$

and

$$\int_0^{2\pi} \chi(t)\, dt = 0.$$

Hence if

$$\chi(x) \sim \sum_{r=-\infty}^{\infty} \chi_r e^{irx}$$

then $\chi_0 = 0$ and

$$\sum_{r=-\infty}^{\infty} |\chi_r|^2$$

converges. By the Fourier series integration theorem and (48)

$$f(x) = \frac{1}{1 - e^{-ix}} \sum_{r=-\infty}^{\infty} \frac{\chi_r}{r} (e^{irx} - 1).$$

Hence

$$f(x)\,(1 - e^{-ix}) = \sum_{r=-\infty}^{\infty} \frac{\chi_r}{r} e^{irx} - \sum_{r=-\infty}^{\infty} \frac{\chi_r}{r} \tag{49}$$

since both of these series converge absolutely.

Now

$$f(x)\,(1 - e^{-ix}) \sim (1 - e^{-ix}) \sum_{n=-\infty}^{\infty} c_n e^{inx}$$

$$= \sum_{n=-\infty}^{\infty} (c_n - c_{n+1})\, e^{inx}.$$

By (49) and the uniqueness theorem for Fourier series of functions of $L^2(0, 2\pi)$ it follows that for $n \neq 0$

$$\frac{\chi_n}{n} = c_n - c_{n+1}.$$

Hence the series

$$\sum_{n=-\infty}^{\infty} n^2 |c_n - c_{n+1}|^2 = \sum_{n=-\infty}^{\infty} |\chi_n|^2$$

converges, and therefore the sequence $\{c_n\}$ belongs to the class $\sum_1^2(-\infty, \infty)$, as required.

By Theorem 9 the series (47) converges for $x \not\equiv 0 \bmod 2\pi$ to a function of $S_1^2(0, 2\pi)$ which must therefore be equal to $f(x)$ almost everywhere. From (34) functions of $S_1^2(0, 2\pi)$ are continuous for $x \not\equiv 0 \bmod 2\pi$, so the sum of the series (47) must be equal to $f(x)$ for all such x.

References

1. A. P. Guinand, *Summation formulae and self-reciprocal functions*, I, II, Quart. J. Math. (1), *9* (1938), 53–67, and (1), *10* (1939), 104–118.
2. ——— *On Poisson's summation formula*, Ann. Math. (2), *42* (1941), 591–603.
3. ——— *General transformations and the Parseval theorem*, Quart. J. Math. (1), *12* (1941), 51–56.
4. G. H. Hardy, J. E. Littlewood, and G. Polya, *Inequalities* (Cambridge, 1934), 239–246.
5. J. B. Miller, *A symmetrical convergence theory for general transforms*, Proc. London Math. Soc. (3), *8* (1958), 224–241.
6. E. C. Titchmarsh, *The Theory of Functions* (Oxford, 1939).
7. ——— *An introduction to the theory of Fourier integrals* (Oxford, 1948).

University of Saskatchewan

THE ANALYTIC CONTINUATION OF THE RIEMANN-LIOUVILLE INTEGRAL IN THE HYPERBOLIC CASE

MARCEL RIESZ

Introduction. In 1949 I published in the *Acta Mathematica* (vol. 81) a rather long paper: "L'intégrale de Riemann-Liouville et le problème de Cauchy." This work will be quoted in the sequel as *Acta* paper. Only minor local references to this paper will be made here, and knowledge of it is not required for the reading of the present article. The notations used here are slightly different from those used in my former paper.

In the *Acta* paper I introduce multiple integrals I^α and I^α_* of the Riemann-Liouville type depending on a parameter α and converging for sufficiently large values of α. I give the solution of the Cauchy problem for the wave equation in a unique formula, the same for space-time of odd or even dimensions, implying an analytic continuation with respect to the parameter α. When this analytic continuation is carried out, it leads to final formulae of quite different types for odd or even dimensions, the one relative to even dimensions obeying the Huygens principle.

The main difficulty concerning the analytic continuation was to prove that I^0 is the identity operator. My way of doing this was neither simple nor elegant. The principal aim of the present paper is to give a more satisfactory proof.

I hope that the present approach will be useful in other connections as well. Indeed, this method of analytic continuation has found unexpected applications in other fields. Here I only make reference to results of Gelfand and Grajew.[1]

1. Preliminaries. If the co-ordinates of a point x in m-dimensional space-time or Lorentz-space are denoted by $x^0, x^1, \dots, x^{m-1}$, the metric form will be

$$(x, x) = (x^0)^2 - (x^1)^2 - \dots - (x^{m-1})^2 = l_{ik}x^ix^k, \tag{1.1}$$

where the ordinary summation convention is used. The square of the distance of two points x and y is given by

$$R_{xy} = r^2_{xy} = (x - y, x - y) = l_{ik}(x^i - y^i)(x^k - y^k). \tag{1.2}$$

The scalar product (a, b) of two vectors a and b, with the respective components a^k and b^k, is defined by

Received October 19, 1959.

[1]See Appendix III of the book by I. M. Gelfand and M. A. Neumark, *Unitäre Darstellungen der klassischen Gruppen* (Berlin: Akademie-Verlag, 1957), also A.M.S. translations, Series 2, vol. 9, pp. 123–154.

$$(a, b) = l_{ik}a^i b^k. \tag{1.3}$$

Two vectors whose scalar product vanishes are said to be *orthogonal* to each other. In what follows, *orthogonality and normality are always meant in this sense.*

According as the scalar square (a, a) of a vector a is (1) positive, (2) zero, (3) negative, the vector is said to be (1) *time-like*, (2) *light-like* or a *null vector*, (3) *space-like*. A time-like or a light-like vector a is called *positive* or *negative* according as its time component a^0 is positive or negative.

Time-like unit vectors u and *space-like unit vectors* v are defined by the relations $(u, u) = 1$ and $(v, v) = -1$ respectively.

The *light cone* or characteristic cone with vertex a is given by the equation $(x - a, x - a) = 0$. The *positive* and *negative* half-cones correspond to $x^0 - a^0 > 0$ or < 0 respectively. These half-cones will be called *positive and negative light cones* in the sequel.

Consider now a p-dimensional (curved) variety S whose points y are referred to p parameters $\lambda^1, \lambda^2, \ldots, \lambda^p$. The p-dimensional *volume element* dS of S, or alternately *surface element* if $1 < p < m$, can be defined in the following way (cf. *Acta* paper pp. 44–45). Let $ds^2 = (dy, dy) = \sum_{i,k} \gamma_{ik} d\lambda^i d\lambda^k$ be the square of the arc element in S. Form the determinant $\gamma = |\gamma_{ik}|$. Then

$$dS = \sqrt{|\gamma|}\, d\lambda^1 d\lambda^2 \ldots d\lambda^p. \tag{1.4}$$

An $(m-1)$-dimensional surface is said to be space-like if its normal is time-like. Let S be a space-like surface. Suppose that the negative light cone C^x with vertex x and the surface S enclose a bounded domain D_S^x. We shall consider functions defined in domains including D_S^x and make the blanket hypothesis that the functions and all their derivatives with respect to the Cartesian co-ordinates which explicitly or implicitly enter into our computations exist and are continuous. We express this by saying that the functions are *well behaved.* The same phrase will be used in an appropriate sense in connection with the surface S and functions defined on S.

We form the *volume potential*

$$I^\alpha f(x) = \frac{1}{H_m(\alpha)} \int_{D_S^x} f(y) r_{xy}^{\alpha-m} dV, \tag{1.5}$$

where $dV = dy^0 dy^1 \ldots dy^{m-1}$ is the volume element of m-space and

$$H_m(\alpha) = \pi^{\frac{1}{2}(m-2)} 2^{\alpha-1} \Gamma(\tfrac{1}{2}\alpha) \Gamma(\tfrac{1}{2}(\alpha + 2 - m)). \tag{1.6}$$

The integral in (1.5) converges for $\alpha > m-2$ (cf. *Acta* paper, p. 31), or more generally for $Re\, \alpha > m$-2, if we admit complex values of α. Similarly, our subsequent assertions about convergence of integrals or analytic continuation of I^α or $I^{*\alpha}$ (see below) remain valid for complex α, if we replace all inequalities of the type $\alpha > \alpha_0$ by $Re\, \alpha > Re\, \alpha_0$.

Besides the volume potentials we also consider potentials of a *simple layer* and of a *double layer*.

Let S^x be that part of the surface S which is interior to the cone C^x. Denote further by n the positive unit normal to S and let g and h be two functions defined on S. We write

$$(1.7)\qquad I_*^\alpha \overline{f, g, h}(x) = \frac{1}{H_m(\alpha)} \int_{D_*^x} f(y) r_{xy}^{\alpha-m} dV + \frac{1}{H_m(\alpha)} \int_{S^x} \left\{ g(y) r_{xy}^{\alpha-m} - h(y) \frac{d}{dn} r_{xy}^{\alpha-m} \right\} dS,$$

where dS is the surface element of S (cf. formula (1.4)).

The simple layer converges for $\alpha > m-2$, while the double layer, whose kernel has a stronger singularity, converges only for $\alpha > m$ (cf. *Acta* paper, pp. 48–49, and §4 of the present paper).

We will show that by virtue of our hypotheses about the behaviour of the surface S and the functions f, g, h the integral I_*^α can be continued analytically down to an arbitrary value $\alpha_0 \gtreqless 0$. Moreover, if $\alpha_0 < 0$, then

$$I_*^0 \overline{f, g, h}(x) = I^0 f(x) = f(x).$$

For a specification of the derivatives needed for different purposes cf. *Acta* paper, pp. 59–60, 64, 223.

Some simple facts concerning the analytic continuation of the ordinary Riemann-Liouville integral in *one dimension* will be needed in the sequel (cf. *Acta* paper, pp. 14–16).

Set

$$(1.9)\qquad J^\alpha f(0) = \frac{1}{\Gamma(\alpha)} \int_0^b f(t) t^{\alpha-1} dt.$$

If $f(t)$ is continuous in the closed interval $[0, b]$, this integral is convergent for $\alpha > 0$. If for $k \leqq n$ the derivatives $f^{(k)}(t)$ exist and are continuous in $[0, b]$, then *$J^\alpha f(0)$ has a holomorphic continuation to all $\alpha > -n$. Moreover, if p is an integer $0 \leqq p < n$, then*

$$(1.10)\qquad J^{-p} f(0) = (-1)^p f^{(p)}(0).$$

(As a matter of fact, only the case $p = 0$ will be used explicitly in the sequel.)

To prove this, set

$$P(t) = \sum_{k=0}^{n-1} \frac{f^{(k)}(0)}{k!} t^k.$$

Then we have for $\alpha > 0$, to begin with, and subsequently by analytic continuation, for all $\alpha > -n$

$$(1.11)\qquad J^\alpha f(0) = \frac{1}{\Gamma(\alpha)} \int_0^b [f(t) - P(t)] t^{\alpha-1} dt + \frac{1}{\Gamma(\alpha)} \sum_{k=0}^{n-1} \frac{f^{(k)}(0)}{k!(k+\alpha)} b^{k+\alpha}.$$

Indeed, the last integral is convergent and the whole expression (1.11) is holomorphic for $\alpha > -n$. $J^\alpha f(0)$ reduces to $(-1)^p f^{(p)}(0)$ at $\alpha = -p$, since $\Gamma(\alpha)$ has a simple pole with residue $(-1)^p/p!$ at this point.

The following extension will also be needed, and the corresponding result will be quoted as *the extended one-dimensional case.* Its verification is left to the reader.

Let $f(t)$ also depend on α. If $f(t)$ and its derivatives with respect to t up to the order n are continuous in the closed interval $[0, b]$ and moreover are holomorphic in α for $\alpha > -n$, then our above statement and its proof remain valid, except for some slight changes in the notations.

2. A co-ordinate system. We place the origin O at the point x and will eventually refer the domain $D_S{}^O$ to co-ordinates which are to be introduced here. We denote a *fixed* negative time-like unit vector by a and a *variable* space-like unit vector orthogonal to a by v. In a suitable Lorentz frame a and v can be written $a = (-1, 0, \ldots, 0)$ and $v = (0, v^1, \ldots, v^{m-1})$ with $\sum (v^k)^2 = 1$. If the vector v issues from the origin, its endpoint describes the unit sphere S_{m-2} lying in the $(m-1)$-plane orthogonal to a. We write out explicitly that

$$(a, a) = 1, \; (v, v) = -1, \; (a, v) = 0. \tag{2.1}$$

An arbitrary position vector y can be written

$$y = ta + \rho v, \; \rho \geqq 0. \tag{2.2}$$

We always suppose that also $t \geqq 0$. This inequality is obviously satisfied in the domain $D_S{}^O$.

The relation (2.2) can also be written

$$y = \tfrac{1}{2}(t + \rho)(a + v) + \tfrac{1}{2}(t - \rho)(a - v).$$

Furthermore, if we set

$$b = \tfrac{1}{2}(a + v), \; c = \tfrac{1}{2}(a - v), \tag{2.3}$$

then

$$y = (t + \rho)b + (t - \rho)c = (t + \rho)\left(b + \frac{t - \rho}{t + \rho}c\right).$$

Setting now

$$\tau = \frac{t - \rho}{t + \rho}, \quad \sigma = t + \rho, \tag{2.4}$$

we obtain

$$y = \sigma(b + \tau c). \tag{2.5}$$

The inverted formulae (2.4) are

$$\rho = \tfrac{1}{2}\sigma(1 - \tau), \; t = \tfrac{1}{2}\sigma(1 + \tau). \tag{2.6}$$

It follows from (2.1) and (2.3) that

$$(b, b) = 0, \; (c, c) = 0, \; (b, c) = \tfrac{1}{2}. \tag{2.7}$$

Hence b and c are (negative) *null vectors.*

The variables τ and σ and the angular variable v which varies on the sphere S_{m-2} and determines the vectors b and c will be our new co-ordinates. Here are the principal merits of τ and σ. The square of the Lorentz distance of a point from the vertex can be expressed and "separated" in τ and σ. The vertex of the cone C^0 is given by the single equation $\sigma = 0$, while the cone apart from the vertex is given by the equation $\tau = 0$. The derivatives $\partial^p f/\partial\tau^p$ of an arbitrary function $f(y)$ vanish at the vertex since they contain the factor σ^p.

We prove these assertions and complete them in certain respects. The square r^2 of the Lorentz distance is according to (2.2) and (2.1)

$$(y, y) = (ta + \rho v, ta + \rho v) = t^2 - \rho^2 = (t + \rho)^2 \frac{t - \rho}{t + \rho}.$$

Hence

$$r^2 = (y, y) = \sigma^2\tau. \tag{2.8}$$

The same relation also follows from (2.5) and (2.7), since (2.7) gives $(b + \tau c, b + \tau c) = \tau$. The equation of the cone C^0 is $(y, y) = 0$. At the vertex $\sigma = 0$, while τ is indeterminate. On the cone, except at the vertex, $\tau = 0, \sigma > 0$. It follows from (2.4) that $0 < \tau \leqq 1$ inside the cone and that, in particular, $\tau = 1$ on the axis $y = \sigma a$ (or $\rho = 0$) and only there.

We always have $\sigma \geqq 0$, according to (2.4) and the inequalities $t \geqq 0$, $\rho \geqq 0$. The equation $\sigma = \text{const.} = \gamma > 0$, which, in view of (2.4), is equivalent to $t + \rho = \gamma$ is the equation of a *positive light cone* $C_{\gamma a}$, with the vertex γa. It is clear that the inequalities $0 \leqq \tau \leqq 1$ and $0 \leqq \sigma \leqq \gamma$ characterize the interior and the boundary of a double cone $D_{\gamma a}{}^0$ limited by the negative light cone C^0 and the positive light cone $C_{\gamma a}$.

From now on we make ample use of our hypothesis that the function $f(y)$ is well behaved (cf. p. 38). We have

$$\frac{\partial y}{\partial t} = \frac{\partial}{\partial \tau}[\sigma(b + \tau c)] = \sigma c, \quad \frac{\partial^p y}{\partial \tau^p} = 0, \quad p = 2, 3, \ldots. \tag{2.9}$$

From this it follows for any function $f(y)$

$$\frac{\partial f}{\partial \tau} = \sigma\Big(\sum c^k \partial_k\Big) f, \quad \text{where } \partial_k = \frac{\partial}{\partial y^k},$$

and, more generally, for any positive integer p,

$$\frac{\partial^p f}{\partial \tau^p} = \sigma^p\Big(\sum c^k \partial_k\Big)^p f. \tag{2.10}$$

This proves our assertion about the behaviour of the derivatives with respect to τ at the vertex.

3. The volume potential. If the surface element of the sphere S_{m-2} is denoted by dS_{m-2}, the volume element dV of the m-space can be written

$dV = \rho^{m-2}d\rho dt dS_{m-2}$. From (2.6) we have that $\rho = \frac{1}{2}\sigma(1-\tau)$ and that the Jacobian

$$\frac{d(\rho, t)}{d(\sigma, \tau)} = \frac{1}{2}\sigma.$$

Making use of (2.5) and (2.8) we obtain after some simplifications

$$(3.1) \qquad \frac{1}{H_m(\alpha)} f(y) r^{\alpha-m} dV$$
$$= \frac{2^{1-m}}{H_m(\alpha)} f[\sigma(b+\tau c)] \sigma^{\alpha-1} \tau^{\frac{1}{2}(\alpha-m)} (1-\tau)^{m-2} d\tau d\sigma dS_{m-2}.$$

In order to get $I^\alpha f(O)$, we have to integrate this expression over the domain $D_S{}^O$. However, it will be convenient to divide this domain into two parts and treat these parts separately. First we choose γ small enough, so that the double-cone $D_{\gamma a}{}^O$ should be contained in $D_S{}^O$. Then we divide the latter domain into $D_{\gamma a}{}^O$ and $D_S{}^O - D_{\gamma a}{}^O$ and show by rather different methods that the corresponding parts of the integral I^α, denoted incidentally by $I_I{}^\alpha$ and $I_{II}{}^\alpha$, are holomorphic for $\alpha > -1$ and that $I_I{}^0 = f(O)$, $I_{II}{}^0 = 0$, which gives that the original $I^0 f(O) = f(O)$. By this our *main objective* will be attained. The more difficult part of the proof, the one concerning I_I, will be carried out in the present section. The easy part I_{II} can be treated by a method similar to that used in §4 for a simple layer. Therefore it is postponed to §5. In the same section we apply our results concerning the analytic continuation of a simple and a double layer to carry out the "unlimited" analytic continuation of the volume potentials.

The integral of the right-hand side of (3.1) extended over the double-cone $D_{\gamma a}{}^O$ gives us the functional $I^\alpha f(O)$ relative to this special domain. A very great simplification arises here from the fact that the limits of integration with respect to τ and σ are fixed, τ varying between 0 and 1 and σ between 0 and γ. Thus we have in the present case

$$I^\alpha f(O) = \int_{S_{m-1}} dS_{m-2} \int_0^\gamma d\sigma \int_0^1 \ldots d\tau,$$

where the dots stand for the integrand given in the right-hand side of (3.1)

Besides the formula (1.6) for $H_m(\alpha)$ we shall need the relations

$$(3.2) \qquad \int_0^1 t^{r-1}(1-t)^{s-1} dt = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)},$$

$$(3.3) \qquad \Gamma(r) = \pi^{-\frac{1}{2}} 2^{r-1} \Gamma(\tfrac{1}{2}r)\Gamma(\tfrac{1}{2}r + \tfrac{1}{2})$$

and the explicit expression for the total surface $|S_{m-2}|$ of the sphere S_{m-2},

$$(3.4) \qquad |S_{m-2}| = \frac{2\pi^{\frac{1}{2}m-\frac{1}{2}}}{\Gamma(\frac{1}{2}m - \frac{1}{2})}.$$

We develop $f[\sigma(b+\tau c)]$ in a finite power series in τ with a remainder term. We have

$$(3.5)\qquad f(y) = f[\sigma(b + \tau c)] = \sum_{p=0}^{N-1} \frac{\tau^p}{p!} \Phi_p(\sigma, v) + R_N(\tau),$$

where $\Phi_p(\sigma, v) = \left\{\frac{\partial^p}{\partial \tau^p} f[\sigma(b + \tau c)]\right\}_{\tau=0}$, and

$$(3.6)\qquad R_N(\tau) = \frac{1}{(N-1)!}\int_0^\tau \frac{\partial^N}{\partial \bar{\tau}^N} f[\sigma(b + \bar{\tau} c)](\tau - \bar{\tau})^{N-1} d\bar{\tau}.$$

N is here a sufficiently large integer, to be specified later. Obviously $R_N(\tau) = O(\tau^N)$.

We first compute the expression

$$A(\alpha) = \frac{2^{1-m}}{H_m(\alpha)}\int_0^1 f[\sigma(b + \tau c)]\tau^{\frac{1}{2}(\alpha-m)}(1-\tau)^{m-2} d\tau.$$

This integral and all the integrals which follow are convergent for $\alpha > m - 2$. On account of (3.5) we have

$$(3.7)\quad A(\alpha) = \sum_{p=0}^{N-1} A_p(\alpha)\, \Phi_p(\sigma, v) + \frac{2^{1-m}}{H_m(\alpha)}\int_0^1 R_N(\tau)\tau^{\frac{1}{2}(\alpha-m)}(1-\tau)^{m-2} d\tau,$$

where by means of (3.2) with $r = \frac{1}{2}(\alpha + 2 - m) + p$, $s = m - 1$

$$(3.8)\qquad A_p(\alpha) = \frac{2^{1-m}\Gamma(m-1)\Gamma(\frac{1}{2}(\alpha + 2 - m) + p)}{p! H_m(\alpha)\Gamma(\frac{1}{2}(\alpha + m) + p)}.$$

The most important term in (3.7) is $A_0(\alpha)\Phi_0(\sigma, v) = A_0(\alpha) f(\sigma b)$. In view of the expression (1.6) of $H_m(\alpha)$ we find

$$(3.9)\qquad A_0(\alpha) = \frac{2^{1-m}\Gamma(m-1)}{\pi^{\frac{1}{2}(m-2)} 2^{\alpha-1}\Gamma(\frac{1}{2}\alpha)\Gamma(\frac{1}{2}(\alpha+m))}.$$

Expressing $2^{\alpha-1}\Gamma(\frac{1}{2}\alpha)$ by means of (3.3), with $r = \alpha$, we find after some simplifications

$$(3.10)\quad A_0(\alpha) = K_0(\alpha)\cdot\frac{1}{\Gamma(\alpha)},\ \text{where } K_0(\alpha) = \frac{2^{1-m}\Gamma(m-1)\Gamma(\frac{1}{2}\alpha + \frac{1}{2})}{\pi^{\frac{1}{2}(m-1)}\Gamma(\frac{1}{2}(\alpha+m))}.$$

Since $\Gamma(\frac{1}{2}) = \pi^{\frac{1}{2}}$, we have

$$(3.11)\qquad K_0(0) = \frac{2^{1-m}\Gamma(m-1)}{\pi^{\frac{1}{2}(m-2)}\Gamma(\frac{1}{2}m)}.$$

According to (3.3), with $r = m - 1$, and to (3.4)

$$(3.12)\qquad K_0(0) = \frac{\Gamma(\frac{1}{2}m - \frac{1}{2})}{2\pi^{(\frac{1}{2}m - \frac{1}{2})}} = \frac{1}{|S_{m-2}|}.$$

Our next step is to carry out the analytic continuation of the expression

$$(3.13)\qquad A_0(\alpha).\int_0^\gamma \Phi_0(\sigma, v)\sigma^{\alpha-1} d\sigma = K_0(\alpha)\cdot\frac{1}{\Gamma(\alpha)}\int_0^\gamma f(\sigma b)\sigma^{\alpha-1} d\sigma.$$

The integral converges for $\alpha > 0$ and, according to what we know about the *one-dimensional case* (cf. p. 39), the analytic continuation of (3.13) is holomorphic for $\alpha > -1$. (For $\alpha = -1$ the function $\Gamma(\frac{1}{2}(\alpha + 1))$ has a pole.) For

$\alpha = 0$ the expression (3.13) becomes $K_0(0)\, f(O) = f(O)/|S_{m-2}|$. The integral of this constant value with respect to the angular variable is clearly $f(O)$. Hence, and this contains virtually our main result, the term corresponding to $p = 0$ in the I^α relative to the double-cone $D_{\gamma a}{}^O$ *yields exactly* $f(O)$ *for* $\alpha = 0$.

The terms in (3.7) with $p > 0$ are easy to handle. In analogy with (3.13) we have to consider the expression

$$(3.14) \qquad A_p(\alpha) \cdot \int_0^\gamma \Phi_p(\sigma, v)\sigma^{\alpha-1} d\sigma, \quad p \geqq 1,$$

where $A_p(\alpha)$ is given in (3.8). We first note that

$$\Gamma(\tfrac{1}{2}(\alpha + 2 - m) + p) = \Gamma(\tfrac{1}{2}(\alpha + 2 - m))P_p(\alpha),$$

where $P_p(\alpha)$ is a polynomial of degree p in α. Hence, after the same simplifications as those performed for $A_0(\alpha)$ we obtain

$$A_p(\alpha) = K_p(\alpha) \cdot \frac{1}{\Gamma(\alpha)}, \text{ where } K_p(\alpha) = \frac{2^{1-m}\Gamma(m-1)\Gamma(\frac{1}{2}(\alpha+1))}{p!\,\pi^{(\frac{1}{2}m-1)}\Gamma(\frac{1}{2}(\alpha+m)+p)} \cdot P_p(\alpha).$$

According to (2.10) $\partial^p/\partial\tau^p$ contains the factor σ^p, $p \geqq 1$. Hence all the integrals of the type given in (3.14) converge if $\alpha > -1$. Moreover, $K_p(0)$ is finite, $1/\Gamma(0) = 0$, consequently all expressions (3.14) vanish for $\alpha = 0$.

Since $R_N = O(\tau^N \sigma^N)$, the remainder term can be treated in an analogous way. It is clear that for the present purposes N may be any integer such that $-\tfrac{1}{2}(1 + m) + N \geqq -1$ or $2N \geqq m - 1$.

Summing up, it is now shown that the integral $I^\alpha f(O)$ extended to the double cone can be analytically continued to all values $\alpha > -1$ and that for these values it is a holomorphic function of α. Moreover $I^0 f(O) = f(O)$, that is I^0 *is the identity operator in the case of the double cone.*

We could have gone a bit farther and established the possibility of the analytic continuation down to arbitrary negative values of α. However, one difficulty would have remained, the possible occurrence of (simple) poles at the negative odd integers = the poles of $\Gamma(\tfrac{1}{2}(\alpha + 1))$. As a matter of fact, none of these poles actually occurs. Their disappearance must be the effect of the integration with respect to the angular variable, considered here in a very summary way. On p. 64 of my *Acta* paper I indicate how the holomorphic character of the unlimited analytic continuation can be established by an indirect method. This will be carried out here in §5.

4. Simple layer and double layer. We now pass to the simple layer

$$(4.1) \qquad \frac{1}{H_m(\alpha)} \int_{S^O} g(y) r^{\alpha-m} dS$$

considered in formula (1.7), where now the vertex coincides with the origin and r is written instead of r_{xy}. The integral converges for $\alpha > m - 2$ and has to be continued analytically for $\alpha \leqq m - 2$. The portion S^O of the surface S can be parametrized by the variables τ and v in the following way. Through

every point of S passes a unique ray issuing from the origin. On such a ray τ and v are constant, hence also the vectors b and c corresponding to v are constant, while σ varies. If we write the point of intersection of the ray with the surface S in the form $y = \sigma_S(b + \tau c)$, the equations $\sigma = \sigma_S(\tau, v)$ or $y = \sigma_S(\tau, v)(b + \tau c)$ and the additional condition $0 \leqq \tau \leqq 1$ yield the required parametrization of S^O, because b and c depend only on v.

Since v is indeterminate on the axis $y = \sigma a$, where $\tau = 1$, we divide S^O into two parts $S_A{}^O$ and $S_B{}^O$ in the following way. With an arbitrary δ such that $0 < \delta < 1$ the first part will be given by $0 \leqq \tau \leqq \delta$ and the second by $\delta < \tau \leqq 1$.

That part of the simple layer which relates to $S_B{}^O$ is an entire function of α vanishing for all even integers $\leqq 0$. Indeed the corresponding part of the integral in (4.1) never ceases to converge and $H_m(\alpha)$ has poles at these integers owing to the factor $\Gamma(\frac{1}{2}\alpha)$ (cf. formula (1.6)).

In order to treat that part of (4.1) which is taken over $S_A{}^O$, we have to express the surface element dS in a convenient way. The angular variable v on the sphere S_{m-2} can be expressed by $m - 2$ local parameters $\phi^1, \phi^2, \ldots,$ ϕ^{m-2}. Thus, according to (1.4), we can write in summary notations

$$dS = G(\tau, v) \,.\, d\tau \,.\, \Pi\, d\phi^i, dS_{m-2} = \Theta(v) \,.\, \Pi\, d\phi^i,$$

hence $dS = H(\tau, v) d\tau dS_{m-2}$.

We set $\frac{1}{2}(\alpha + 2 - m) = \beta$ and can then write according to (1.6)

$$H_m(\alpha) = L_m(\alpha)\Gamma(\beta) \text{ where } L_m(\alpha) = \pi^{\frac{1}{2}m-1}2^{\alpha-1}\Gamma(\tfrac{1}{2}\alpha). \tag{4.2}$$

We also set $g(y) = g(\tau, v)$ and recall the relation $r^2 = \sigma^2\tau$ given in (2.8). Then we write that part of (4.1) which corresponds to $S_A{}^O$ in the form

$$U(\alpha) = \frac{1}{L_m(\alpha)} \int_{S_{m-2}} dS_{m-2} \frac{1}{\Gamma(\beta)} \int_0^\delta g(\tau, v) H(\tau, v)[\sigma_S(\tau, v)]^{\alpha-m} \tau^{\beta-1} d\tau. \tag{4.3}$$

Here $H(\tau, v)$, $\sigma_S(\tau, v)$, $g(\tau, v)$ are well-behaved even in τ and v by virtue of our hypothesis concerning the surface S and the function $g(y)$. Moreover σ_S is bounded away from 0. Hence we can apply our statement concerning the *extended one-dimensional case* (cf. p. 40), which gives that $(1/(\Gamma(\beta))\int_0^\delta \ldots$ is a holomorphic function of β, hence also of α, and this is then also true for $U(\alpha)$. Owing to the presence of the factor $\Gamma(\frac{1}{2}\alpha)$ in $L_m(\alpha)$, the function $U(\alpha)$ vanishes for $\alpha = 0$ and $\alpha =$ a negative even integer.

There is very little to change in the case of a double layer. It is easily seen that $dr/dn = r^{-1}(y, n)$ hence

$$\frac{dr^{\alpha-m}}{dn} = (\alpha - m)r^{\alpha-m-1}\frac{dr}{dn} = (\alpha - m)r^{\alpha-m-2}(y, n).$$

The scalar product (y, n) (cf. (1.3)) is a well behaved function in τ and v. Owing to the lowered exponent the double layer integral in (1.7) converges only for $\alpha > m$, thus the need of continuation begins already at m.

The part relative to $S_B{}^O$ is again an entire function of α and vanishes for all even integers $\leqq 0$. On the other hand, with $\beta' = \frac{1}{2}(\alpha - m) = \beta - 1$,

$$\frac{\alpha - m}{\Gamma(\beta)} r^{\alpha-m} = \frac{2(\beta - 1)}{\Gamma(\beta)} r^{\beta-1} = \frac{2}{\Gamma(\beta - 1)} r^{\beta-1} = \frac{2}{\Gamma(\beta')} r^{\beta'}.$$

Thus, when treating the part relative to $S_A{}^O$, we obtain a formula of the same type as (4.3), $\beta' = \beta - 1$ playing for the double layer the same role as β played for the simple layer, and the results are essentially the same.

Our findings can be summed up as follows. *Both the simple layer and the double layer potentials can be continued analytically to arbitrary values of α. They are holomorphic functions of α which vanish for all even integers $\leqq 0$.*

5. The volume potential (continued). Now we return to the volume potential and clarify the properties of the part which relates to the domain $D_S{}^O - D_{\gamma_s}{}^O$ (cf. the beginning of §3). We want to prove that this part of I^α, when continued analytically, is holomorphic for $\alpha > -1$, and vanishes for $\alpha = 0$.

In the same way as we did in the previous section with the portion of surface S^O, we now divide the domain $D_S{}^O - D_{\gamma_s}{}^O$ into two parts according as $0 \leqq \tau \leqq \delta$ or $\delta < \tau \leqq 1$. The volume potential relative to the second part is again an entire function vanishing for all even integers $\leqq 0$. Thus we only have to investigate the first part. This can be written in the form (cf. (3.1.))

$$\frac{2^{1-m}}{H_m(\alpha)} \int_{S_{m-2}} dS_{m-2} \int_0^\delta \tau^{\frac{1}{2}(\alpha-m)} (1 - \tau)^{m-2} d\tau \int_\gamma^{\sigma_S} f[\sigma(b + \tau c)] \sigma^{\alpha-1} d\sigma,$$

where σ_S, defined in the previous section, depends on τ and v, $\sigma_S = \sigma_S(\tau, v)$. We set

$$\int_\gamma^{\sigma_S} f[\sigma(b + \tau c)] \sigma^{\alpha-1} d\sigma = F(\tau, v, \alpha),$$

where F is well behaved in τ (and v) and holomorphic in α, since σ is bounded away from 0. Writing, as in the case of a simple layer, $H_m(\alpha) = L_m(\alpha) \Gamma(\beta)$ we see by virtue of our findings in the extended one-dimensional case that

$$\frac{1}{L_m(\alpha)} \frac{1}{\Gamma(\beta)} \int_0^\delta F(\tau, v, \alpha) \tau^{\beta-1} (1 - \tau)^{m-2} d\tau$$

can be continued analytically as a holomorphic function of α to any $\alpha > \alpha_0$, where α_0 is arbitrary. Thus it is a holomorphic function for $\alpha > -1$ any way, and again, owing to the presence of the factor $\Gamma(\alpha/2)$ in $L_m(\alpha)$, it vanishes for $\alpha = 0$.

This completes the proof of the fact stated in §3, that $I^0 f(O) = f(O)$, if I^0 is relative to the *original domain* $D_S{}^O$. This can clearly be expressed by the more inspiring formula

(5.1) $$I^0 f(x) = f(x),$$

or by the statement that I^0 *is the identity operator*. This is *our principal result.* The passage from (5.1) to the relation

$$I^0_\bullet \overline{f, g, h}\,(x) = f(x) \tag{5.2}$$

follows from the properties of simple and double layers established in the previous section.

We conclude this paper by two additional remarks, the first concerning the unlimited analytic continuation of the volume potential, the second concerning the case of the infinite cone.

As indicated on p. 64 of our *Acta* paper, the unlimited holomorphic continuation of the volume potential can be reduced to that of simple layer and double layer potentials.

We consider the wave operator

$$\Delta = \frac{\partial^2}{(\partial x^0)^2} - \frac{\partial^2}{(\partial x^1)^2} - \dots - \frac{\partial^2}{(\partial x^{m-1})^2}.$$

Then, by virtue of Green's formula, we have in the notation (1.7) for $\alpha > m - 2$

$$I^\alpha f(x) = I^{\alpha+2}_\bullet \overline{\Delta f, \frac{df}{dn}, f(x)} \tag{5.3}$$

(see *Acta* paper, pp. 46–47). The left-hand side converges for $\alpha > m - 2$. On the right-hand side the volume potential and the simple layer converge for $\alpha + 2 > m - 2$, that is for $\alpha > m - 4$, while the double layer converges only for $\alpha + 2 > m$, that is only for $\alpha > m - 2$. Thus, seemingly nothing is gained as far as the analytic continuation of $I^\alpha f(x)$ is concerned. But if we take into account our results about the unlimited holomorphic continuation of the simple and the double layer and the fact that $I^{\alpha+2}\Delta$ is holomorphic for $\alpha > m - 4$, the possibility of the holomorphic continuation of $I^\alpha f(x)$ down to $m - 4$ is established. The iteration of this procedure, that is the application of formula (5.3) to $I^{\alpha+2}\Delta f$, $I^{\alpha+4}\Delta^2 f, \dots$, establishes the possibility of an unlimited holomorphic continuation of $I^\alpha f(x)$.

In the case of an infinite cone we suppose that $f(x)$ and its derivatives are not only well behaved, but also decrease rapidly enough at infinity. In this case formula (5.3) reduces to

$$I^\alpha f(x) = I^{\alpha+2}\Delta f(x), \tag{5.4}$$

and the iteration of this formula gives immediately the possibility of an unlimited holomorphic continuation. However, in order to establish the main relation $I^0 f(x) = f(x)$, we still have to go back to §3 and use the double-cone $D_{\gamma\alpha}{}^O$. The treatment of the complementary expression $I_{II}{}^\alpha$ is in the infinite case still simpler than in the finite case.

University of Lund
University of Maryland

CONSTRUCTION OF PRIMITIVES OF GENERALIZED DERIVATIVES WITH APPLICATIONS TO TRIGONOMETRIC SERIES

P. S. BULLEN

1. Introduction. This paper is an extension of the ideas discussed in **(3**, §§ 14-16**)**; the extension consisting of the use of the third and fourth symmetric Riemann derivative instead of the Schwarz or second symmetric Riemann derivative.

The J_2-integral, due to James **(1)**, is defined in **(3)** as follows. Let $f(x)$ be measurable on $[a, b]$ and finite at each point; if there exists a continuous function $F(x)$ such that $D^2F = f$ everywhere on (a, b),

$$D^2F = \lim_{h\to 0}\frac{F(x+h) - 2F(x) + F(x-h)}{h^2}$$

then

$$\int_{a,b}^{x} f(t)\,d_2 t = F(x) - \frac{x-b}{a-b}F(a) - \frac{x-a}{b-a}F(b) = H_2(F: a, b, x). \tag{1.1}$$

The definition is unique since if $F(x)$ and $G(x)$ are continuous and $D^2F = D^2G$ everywhere then

$$H_2(F: a, b, x) = H_2(G: a, b, x).$$

This integral has application to convergent trigonometric series, **(3)**.

Using the third and fourth symmetric Riemann derivatives J_3- and J_4-integrals are defined and applied to $(C, 1)$ and $(C, 2)$ summable trigonometric series.

2. Definitions. With the notation of Kassimatis, **(4)**, we write for any function $F(x)$ defined at the points x_1, x_2, x_3, x_4,

$$H_3(F: x_1, x_2, x_3, x_4) = F(x_4) - F(x_3)\frac{(x_4 - x_1)(x_4 - x_2)}{(x_3 - x_1)(x_3 - x_2)} - F(x_2)\frac{(x_4 - x_3)(x_4 - x_1)}{(x_2 - x_3)(x_2 - x_1)} - F(x_1)\frac{(x_4 - x_2)(x_4 - x_3)}{(x_1 - x_2)(x_1 - x_3)}, \tag{2.1}$$

$$V_3(F: x_1, x_2, x_3, x_4) = \frac{H_3(F: x_1, x_2, x_3, x_4)}{(x_4 - x_1)(x_4 - x_2)(x_4 - x_3)}. \tag{2.2}$$

Received November 6, 1959. This paper was written while the author was a fellow at the Summer Research Institute of the Canadian Mathematical Congress.

V_3 is then the third divided difference of $F(x)$. In particular if $h > k > 0$ we write

$$(2.3) \quad w_3(F: x; h, k) = w_3(x; h, k) = 3! V_3(F: x + h, x + k, x - k, x - h)$$
$$= \frac{3}{h^2 - k^2} \left\{ \frac{F(x+h) - F(x-h)}{h} - \frac{F(x+k) - F(x-k)}{k} \right\},$$

$$(2.4) \quad w_3(F: x; 3h, h)$$
$$= \frac{\Delta^3(F: 2h)}{(2h)^3} = \frac{F(x+3h) - 3F(x+h) + 3F(x-h) - F(x-3h)}{(2h)^3}.$$

From (2.3) and (2.4) we define

$$(2.5) \qquad \Delta'''F(x) = \overline{\lim_{h,k\to 0}}\, w_3(x; h, k), \qquad \delta'''F(x) = \underline{\lim}_{h,k\to 0}\, w_3(x: h, k),$$

$$(2.6) \qquad \overline{D}^3 F(x) = \overline{\lim_{h\to 0}}\, w_3(x: 3h, h), \qquad \underline{D}^3 F(x) = \underline{\lim}_{h\to 0}\, w_3(x: 3h, h),$$

and if $\underline{D}^3 F(x) = \overline{D}^3 F(x)$ we say that $F(x)$ has a third symmetric Riemann derivative at x and write it $D^3 F(x)$.

Clearly

$$(2.7) \qquad \delta''' F \leq \underline{D}^3 F \leq \overline{D}^3 F \leq \Delta''' F.$$

The following lemma, which generalizes Theorem 19, **(3)**, is needed later.

LEMMA 2.1. *If F'' exists in an interval containing x and if $\Delta_1(\delta_1)$ is the greater (smaller) of the first derivates of F'' then*

$$(2.8) \qquad \delta_1 \leq \delta''' \leq \Delta''' \leq \Delta_1.$$

All points will be assumed to be interior to the interval mentioned in the statement of the lemma. It is sufficient to prove $\delta_1 \leq \delta$ as a similar argument will complete (2.8). Further we may obviously assume $\delta_1 > -\infty$. The proof is in two parts.

(a) Assume $\delta_1 < \infty$. From the definition of δ_1, if $\epsilon > 0$ is given, there exists $\mu > 0$ such that if $0 < \eta, \xi < \mu$ then

$$F''(x + \eta) - F''(x) > \eta(\delta_1 - \epsilon),$$
$$F''(x - \xi) - F''(x) < \xi(\delta_1 - \epsilon).$$

Consider the function $X(u)$ defined by

$$X(u) = F(x + u) - F(x) - uF'(x) - \frac{u^2}{2!} F''(x) - \frac{u^3}{3!}(\delta_1 - \epsilon).$$

The following properties of $X(u)$ are immediate,

$$X'(u) = F'(x + u) - F'(x) - uF''(x) - \frac{u}{2!}(\delta_1 - \epsilon),$$

$$X''(u) = F''(x + u) - F''(x) - u(\delta_1 - \epsilon),$$
$$X(0) = X'(0) = X''(0) = 0,$$

$$(2.9)\qquad \begin{aligned} X''(u) &> 0 \text{ if } 0 < u < \mu, \\ X''(u) &< 0 \text{ if } -\mu < u < 0, \end{aligned}$$

$$(2.10)\qquad w_3(X:0;h,k) = w_3(F:x;h,k) - (\delta_1 - \epsilon).$$

It follows, from (2.10), that it is sufficient to show that, for all h, k small enough, $w_3(X:0;h,k) \geqslant 0$. To do this define, for $0 < u < \mu$,

$$Y(u) = \frac{X(u) - X(-u)}{u}.$$

Then by (2.3)

$$w_3(X:0;h,k) = \frac{3}{h^2 - k^2}(Y(h) - Y(k)).$$

If, therefore, $Y(u)$ is monotonic increasing for all u small enough, then $w_3(X:0;h,k) \geqslant 0$. Now

$$\begin{aligned} Y'(u) &= -\frac{1}{u^2}[\{X(u) - uX'(u)\} - \{X(-u) - (-u)X'(-u)\}] \\ &= -\frac{1}{u^2}[Z(u) - Z(-u)] \end{aligned}$$

where $Z(u) = X(u) - uX'(u)$. $Z(u)$ is clearly defined wherever $X(u)$ and $X'(u)$ are defined and

$$Z'(u) = -uX''(u) \leqslant 0$$

by (2.9). Hence $Z(-u) \leqslant Z(u)$ and hence $Y'(u) \geqslant 0$ wherever $Y'(u)$ is defined.

Thus we have shown that $Y(u)$ is monotonic increasing and the result follows.

(b) Assume $\delta_1 = \infty$. Then in the above argument replace $\delta_1 - \epsilon$ by an arbitrary positive number A to arrive at $\delta''' = \delta_1 = \infty$.

The following lemma due to Saks, **(5)**, will be required later.

LEMMA 2.2. *If $F'(x)$ exists everywhere in $[a, b]$ and $\underline{D}^3F > 0$ in (a, b) then $F'(x)$ is continuous, convex, and*

$$(2.11)\qquad \Delta^3F(x:2h) \geqslant 0$$

for every x, $h > 0$ ($a \leqslant x - 3h < x + 3h \leqslant b$).

3. We now wish to define a class of funtions for which $D^3F = D^3G$ everywhere in an interval implies $H_3(F: x_1, x_2, x_3, x_4) = H_3(g: x_1, x_2, x_3, x_4)$ for all sets of four points in that interval. As has been pointed out by Kassimatis, **(4)**, continuity of F and G is not enough.

LEMMA 3.1. *If $F'(x)$ exists and is continuous in $[a, b]$ then*

$$(3.1)\qquad \min_{a<x<b} \underline{D}^3F(x) \leqslant 3!\, V_3(F: x_1, x_2, x_3, x_4) \leqslant \max_{a<x<b} \overline{D}^3F(x)$$

for all x_1, x_2, x_3, x_4 in $[a, b]$.

The argument is that of Verblunsky, **(6)**. Define

$$f(x) = F(x) - (ax^3 + bx^2 + cx + d) \tag{3.2}$$

where a, b, c, d are determined by the conditions

$$f(x_1) = f(x_2) = f(x_3) = f(x_4) = 0$$

for some x_1, x_2, x_3, x_4 in (a, b). Simple calculations then show that

$$a = V_3(F; x_1, x_2, x_3, x_4).$$

Since $f(x)$ has four zeros, $f'(x)$ has three zeros and a maximum at some point ξ, say. Then we must have

$$w_3(f: \xi: 3h_i, h_i) < 0 \tag{3.3}$$

for a sequence of $h_i, h_i \to 0$. For if not, then for all h small enough

$$\Delta^3 \frac{(f: \xi: 2h)}{(2h)^3} > 0$$

that is,

$$\frac{f(\xi + 3h) - f(\xi - 3h)}{6h} > \frac{f(\xi + h) - f(\xi - h)}{2h}$$
$$> \dots > f'(\xi),$$

which contradicts the fact that $f'(x)$ has a maximum at ξ. As we have (3.3) it follows that

$$\underline{D}^3 f(\xi) < 0$$

that is,

$$\underline{D}^3 F(\xi) < 3!\, a = 3!\, V_3(F: x_1, x_2, x_3, x_4).$$

This proves the left-hand inequality of (3.1). The right-hand inequality comes from applying the above result to $-F$. Finally the result holds for x_1, x_2, x_3, x_4, in $[a, b]$ by continuity of $F(x)$.

An immediate corollary of Lemma 3.1 is

LEMMA 3.2. *If the relation* (3.1) *holds for* $F(x) - G(x)$, *in particular if* $(F - G)'$ *is continuous, then* $D^3 (F - G) = 0$ *implies*

$$H_3(F: x_1, x_2, x_3, x_4) = H_3(G: x_1, x_2, x_3, x_4) \tag{3.4}$$

for all x_1, x_2, x_3, x_4 *in* $[a, b]$.

Let

$$F_1(x) = H_3(F: x_1, x_2, x_3, x)$$
$$G_1(x) = H_3(G: x_1, x_2, x_3, x).$$

Then $D^3(F_1 - G_1) = 0$ and hence, by (3.1),

$$V_3(F_1 - G_1: y_1, y_2, y_3, y_4) = 0$$

for all y_1, y_2, y_3, y_4. $F_1(x) - G_1(x)$ is therefore a polynomial of degree at most 2 but it is zero at x_1, x_2, x_3 and hence is identically zero.

The following lemma, due to Kassimatis, **(4)**, obtains (3.4) under weaker conditions than the continuity of $(F - G)'$ but it is quite possible that (3.1) holds under less restrictive conditions which would then generalize Lemma 3.2.

LEMMA 3.3. *If $F(x)$ and $G(x)$ are defined in $[a, b]$ and* (i) *$F - G$ is continuous in $[a, b]$,* (ii) *$(F - G)'$ exists in (a, b) then $D^3(F - G) = 0$ implies* (3.4).

4. The J_3-integral. Let $f(x)$ be defined and measurable on $[a, b]$ and finite at each point. If there exists a function $F(x)$, continuous on $[a, b]$ and differentiable on (a, b) such that $D^3F = f$ then we define the J_3-integral of f to be

$$\int_{x_1, x_2, x_3}^{x} f(t)d_3t = H_3(F: x_1, x_2, x_3, x) \tag{4.1}$$

where x_1, x_2, x_3, x are any four points of $[a, b]$ and $a \leqslant x_1 < x_2 < x_3 \leqslant b$. Lemma 3.3 ensures that this definition is unique.

If $f(x)$ is complex valued and $f(x) = u(x) + iv(x)$ then we define the J_3-integral of $f(x)$, if it is defined for both $u(x)$ and $v(x)$, by

$$\int_{x_1, x_2, x_3}^{x} f(t)d_3t = \int_{x_1, x_2, x_3}^{x} u(t)d_3 + i \int_{x_1, x_2, x_3}^{x} v(t)d_3(t).$$

The following elementary properties of this integral are immediate.

(a) If $f(x)$ and $g(x)$ are J_3-integrable on $[a, b]$ so is $\alpha f(x) + \beta g(x)$ for any numbers α, β and

$$\int_{x_1, x_2, x_3}^{x} \{\alpha f(t) + \beta g(t)\}d_3t = \alpha \int_{x_1, x_2, x_3}^{x} f(t)d_3t + \beta \int_{x_1, x_2, x_3}^{x} g(t)d_3t.$$

(b) If $f(x)$ is J_3-integrable on $[a, b]$ it is also J_3-integrable on any subinterval $[\alpha, \beta]$ and if

$$F(x) = \int_{a, \gamma, b}^{x} f(t)d_3t \quad \text{then if } \alpha < \delta < \beta \quad \text{and} \quad \alpha < x < \beta$$

$$\int_{\alpha, \delta, \beta}^{x} f(t)d_3t = H_3(F: \alpha, \delta, \beta, x).$$

5. Application to trigonometric series.

THEOREM 1. *Let $f(t)$ be the $(C, 1)$ sum of the series*

$$\sum_{-\infty}^{\infty} c_n e^{int}, \; c_0 = 0,$$

and let

$$F(t) = \sum_{-\infty}^{\infty} \frac{c_n e^{int}}{(in)^3}, \tag{5.1}$$

then

$$\int_{x_1,x_2,x_3}^{x} f(t)d_3t = H_3(F: x_1, x_2, x_3, x). \tag{5.2}$$

To obtain this result we need the following lemma (**6**, II, p. 69).

LEMMA 5.1. "*If*

$$\sum_{-\infty}^{\infty} c_n e^{inx}, \quad c_0 = 0,$$

is summable (C, α), $\alpha > -1$ *at* x_0 *to* s *then it is summable* R_r *at* x_0 *to* s *provided* $r > 1 + \alpha$. *By this we mean that*

$$\lim_{h\to 0} \sum_{-\infty}^{\infty} c_n e^{inx}\left(\frac{\sin n\, h}{inh}\right)^r = s."$$

The result as stated in (**8**) requires $\alpha > -1$ but it is in fact true when $\alpha = -1$ when it is the result of (**8**, I, p. 322).
Simple calculations give

$$\frac{F(x+h) - F(x-h)}{2h} = \sum_{-\infty}^{\infty} \frac{c_n}{in^2} e^{inx}\left(\frac{\sin n h}{inh}\right)$$

$$\frac{F(x+2h) - 2F(x) + F(x-2h)}{(2h)^2} = \sum_{-\infty}^{\infty} \frac{c_n}{i^2 n} e^{inx}\left(\frac{\sin n h}{inh}\right)^2$$

$$\frac{F(x+3h) - 3F(x+h) + 3F(x-h) - 3F(x+3h)}{(2h)^3} = \sum_{-\infty}^{\infty} \frac{c_n}{i} e^{inx}\left(\frac{\sin n h}{inh}\right)^3.$$

By hypothesis the series $\sum c_n e^{int}$ is summable $(C, 1)$ and hence the series

$$\sum \frac{c_n}{n} e^{inx} \quad \text{and} \quad \sum \frac{c_n}{n^2} e^{inx}$$

are summable $(C, 0)$ and $(C, -1)$ respectively.

Hence by Lemma 5.1 $D^3F = f$ everywhere and D^2F and $D\ F$ exist and this implies the existence of $F'(x)$, (4). This then proves that f is J_3-integrable and gives (5.2).

THEOREM 2. *If*

$$\sum_{-\infty}^{\infty} c_n e^{int}$$

is summable $(C, 1)$ *to* $f(t)$ *then*

$$c_n = -\frac{3}{8\pi^3} \int_{-4\pi,-2\pi,2\pi}^{0} f(t) e^{-int}\, d_3t. \tag{5.3}$$

We first calculate c_0. In Theorem 1 we assumed for simplicity that $c_0 = 0$ but this clearly involves no loss in generality.

Hence we know that

$$\int_{-4\pi,-2\pi,2\pi}^{0} f(t)d_3 t = H_3(F: -4\pi, -2\pi, 2\pi, 0)$$

$$= F(0) + \frac{1}{3} F(-4\pi) - F(-2\pi) - \frac{1}{3} F(2\pi),$$

where

$$F(x) = \frac{c_0 x^3}{3!} + \sum_{-\infty}^{\infty}{}' \frac{c_n e^{inx}}{(in)^3}.$$

Since the last term on the right-hand side is periodic its contribution to the integral is zero. Therefore,

$$\int_{-4\pi,-2\pi,2\pi}^{0} f(t)d_3 t = \frac{1}{3}\frac{c_0}{6}(-4\pi)^3 - \frac{c_0}{6}(-2\pi)^3 - \frac{1}{3}\frac{c_0}{6}(2\pi)^3.$$

$$= -\frac{8\pi^3}{3} c_0.$$

To calculate c_n, $n > 0$, requires $f(x)e^{inx}$ to be expressed as the $(C, 1)$ sum of a trigonometric series with constant term c_n. This has been done by James, **(2)**, and then a similar calculation to the one above completes the proof of (5.3).

6. Construction of the J_3-integral. The J_3-integral can be constructed by methods used in Jeffery, **(3)**, to construct the J_2-integral.

THEOREM 3. *Let $f(x) \in L(a, b)$ and let $f(x)$ be the finite third symmetric Riemann derivative of a function continuous in $[a, b]$ and differentiable in (a, b). Further let*

$$\Phi(x) = \int_a^x \int_a^u \int_a^v f(t)\, dt\, dv\, du, \tag{6.1}$$

then

$$\int_{x_1,x_2,x_3}^{x} f(t)d_3 t = H_3(\Phi: x_1, x_2, x_3, x). \tag{6.2}$$

Since the construction follows the lines of **(3)** it is only sketched here to point out certain differences.

We first determine a sequence of continuous functions $U_n(x)$ such that $\underline{D}^3 U_n(x) > f(x)$ and which converges uniformly to $\Phi(x)$.

As in **(1)** define $A_n(x)$ such that, with the notation of Lemma 2.1, $\delta_1 A_n(x) >$ (x) and $A_n(x)$ converges uniformly to $\int_a^x f(t)\, dt$ as $n \to \infty$.

Then the required $U_n(x)$ is

$$U_n(x) = \int_a^x du \int_a^u A_n(t)\, dt,$$

which clearly converges uniformly to $\Phi(x)$ and, from the continuity of $A_n(x)$,

$$U_n'(x) = \int_0^x A_n(t)\, dt, \qquad U_n''(x) = A_n(x).$$

By Lemma 2.1

$$\delta''' U_n(x) \geqslant \delta_1 U_n = \delta_1 A_n > f(x) = D^3 F(x)$$

where $F(x)$ is some function continuous on $[a, b]$ and differentiable on (a, b). Hence

$$\underline{D}^3(U_n(x) - F(x)) > 0$$

and so, by Lemma 2.2,

$$\Delta^3(U_n - f)(x : 2h) > 0$$

for all $x, h > 0, a < x - 3h < x + 3h < b$

Hence letting $n \in \infty$

$$\Delta^3(\Phi - F)(x\ 2h) \geqslant 0$$

which implies

$$\underline{D}^3(\Phi - F) \geqslant 0.$$

In a similar manner it can be shown that

$$\bar{D}^3(\Phi - F) \leqslant 0$$

which together with the previous inequality implies

$$D^3(\Phi - F) = 0.$$

From Lemma 3.2 this gives

$$H_3(F: x_1, x_2, x_3, x_4) = H_3(\Phi: x_1, x_2, x_3, x_4),$$

completing the proof of the theorem.

A function is said to be Lebesgue integrable at a point x_0 if it is Lebesgue integrable in every sufficiently small neighbourhood of x_0.

As in **(3)** the above result can be extended to functions $f(x)$ which have a finite number of points at which they are not Lebesgue integrable. This can be done provided only that if β is such a point, then

$$\lambda(x) = \int_a^x \int_a^u f(t)\, dt\, du$$

is Denjoy integrable in some interval (α, γ) containing β.

7. The fourth symmetric derivative. We now indicate the definitions and results in the case of the fourth symmetric Riemann derivative. As in § 2 we define

$$\text{(7.1)}\quad H_4(F: x_1, x_2, x_3, x_4, x_5) = F(x_5) - F(x_4)\frac{(x_5 - x_1)(x_5 - x_2)(x_5 - x_3)}{(x_4 - x_1)(x_4 - x_2)(x_4 - x_3)}$$
$$- F(x_3)\frac{(x_5 - x_4)(x_5 - x_1)(x_5 - x_2)}{(x_3 - x_4)(x_3 - x_1)(x_3 - x_2)}$$
$$- F(x_2)\frac{(x_5 - x_3)(x_5 - x_4)(x_5 - x_1)}{(x_2 - x_3)(x_2 - x_4)(x_2 - x_1)} - F(x_1)\frac{(x_5 - x_2)(x_5 - x_3)(x_5 - x_4)}{(x_1 - x_2)(x_1 - x_3)(x_1 - x_4)},$$

$$(7.2)\quad V_4(F: x_1, x_2, x_3, x_4, x_5) = \frac{H_4(F: x_1, x_2, x_3, x_4, x_5)}{(x_5 - x_1)(x_5 - x_2)(x_5 - x_3)(x_5 - x_4)}.$$

It may be noted in passing that the function $f(x)$ in (3.2) is equal to $H_4(F: x_1, x_2, x_3, x_4, x)$.

In particular if $h > k > 0$ we write

$$(7.3)\quad w_4(F: x; h, k) = 4!\, V_4(F: x + h, x + k, x - k, x - h)$$

$$= \frac{12}{h^2 - k^2}\left\{\frac{F(x + h) - 2F(x) + F(x - h)}{h^2} - \frac{F(x + k) - 2F(x) + F(x - k)}{k^2}\right\},$$

$$(7.4)\quad \frac{\Delta^4(F: x; h)}{h^4} = w_4(F: x; 2h, h)$$

$$= \frac{F(x + 2h) - 4F(x + h) + 6F(x) - 4F(x - h) + F(x - 2h)}{h^4}.$$

Using (7.3) and (7.4) we define

$$(7.5)\quad \Delta^{(iv)}F = \overline{\lim_{h,k\to 0}}\, w_4(h, k),\ \delta^{(iv)}F = \underline{\lim_{h,k\to 0}}\, w_4(h, k)$$

$$(7.6)\quad \overline{D}^4F = \overline{\lim_{h\to 0}}\, w_4(2h, h),\ \underline{D}^4F = \underline{\lim_{h\to 0}}\, w_4(2h, h),$$

and if $\overline{D}^4F = \underline{D}^4F$ we say that F has a fourth symmetric Riemann derivative and write it D^4F. Clearly

$$(7.7)\quad \delta^{(iv)}F \leqslant \underline{D}^4F \leqslant \overline{D}^4F \leqslant \Delta^{(iv)}F,$$

and we have the following lemmas.

LEMMA 7.1. *If $F'''(x)$ exists in an interval containing x and if $\Delta_1(\delta_1)$ is the greater (smaller) of the first derivates of F then*

$$(7.8)\quad \delta_1 \leqslant \delta^{(iv)} \leqslant \Delta^{(iv)} \leqslant \Delta_1.$$

LEMMA 7.2. *If $F''(x)$ exists everywhere in $[a, b]$ and $\underline{D}^4F \geqslant 0$ in (a, b) then $F''(x)$ is continuous, convex, and*

$$(7.9)\quad \Delta^4F(x: h) \geqslant 0$$

for every x, $h > 0$ $(a \leqslant x - 4h < x + 4h \leqslant b)$.

The proof of Lemma 7.1 is very similar to that of Lemma 2.1. Making all the obvious changes define

$$X(u) = F(x + u) - F(x) - uF'(x) - \frac{u^2}{2!}F''(x) - \frac{u^3}{3!}F'''(x) - \frac{u^4}{4!}(\delta_1 - \epsilon).$$

As in the previous proof it is sufficient to show that $w_4(X: 0; h, k) \geqslant 0$ for all h, k small enough.

Defining

$$Y(u) = \frac{X(u) - 2X(0) + X(-u)}{u^2} = \frac{X(u) + X(-u)}{u^2}$$

it is sufficient to prove $Y(u)$ to be monotonic for u small enough, $u > 0$. Then if we define

$$Z(u) = 2X(u) - uX'(u)$$

it is sufficient to show that $Z(u)$ has a local maximum at $u = 0$. This follows since $Z'(u)$ is monotonic decreasing, $Z''(u)$ being $-uX'''(u)$ which by a result similar to (2.9) is always negative.

The proof of Lemma 7.2 is exactly similar to that of Lemma 2.2 owing to the reasons given by Verblunsky in **(7)**.

LEMMA 7.3. *If $F''(x)$ exists and is continuous in $[a, b]$ then*

$$\min_{a<x<b} \underline{D}^4F(x) < 4'\ V_4(F: x_1, x_2, x_3, x_4, x_5) \leqslant \max_{a<x<b} \overline{D}^4F(x) \tag{7.10}$$

for all x_1, x_2, x_3, x_4, x_5 in $[a, b]$.

LEMMA 7.4. *If (7.10) holds for $F(x) - G(x)$, in particular if $(F - G)''$ is continuous, then $D^4(F - G) = 0$ implies*

$$H_4(F: x_1, x_2, x_3, x_4, x_5) = H_4(G: x_1, x_2, x_3, x_4, x_5). \tag{7.11}$$

LEMMA 7.5. *If $F(x)$ and $G(x)$ are defined in $[a, b]$ and* (i) *$(F - G)$ is continuous in $[a, b]$,* (ii) *$(F - G)''$ exists in (a, b) then $D^4(F - G) = 0$ implies* (7.11).

As the proofs of the corresponding lemmas 3.1, 3.2, and 2.3 depend on **(6)** the proof of these are exactly the same but are based on **(7)**.

Now let $f(x)$ be defined at each point and measurable. If there exists a function $F(x)$, continuous on $[a, b]$ and with a second derivative on (a, b) such that $D^4F = f$, we define the J_4-integral of f to be

$$\int_{x_1,x_2,x_3,x_4}^{x} f(t)d_4t = H_4(F: x_1, x_2, x_3, x_4, x) \tag{7.12}$$

where x_1, x_2, x_3, x_4, x are any five points of $[a, b]$ and $a \leqslant x_1 < x_2 < x_3 < x_4 \leqslant b$. The discussion of § 4 applies with obvious changes to this definition. Further, the following theorems can be proved.

THEOREM 4. *If*

$$\sum_{-\infty}^{\infty} c_n e^{inx}, c_0 = 0,$$

is $(C, 2)$ summable everywhere to $f(t)$ and $c_n = o(n)$ and if

$$F(t) = \sum \frac{c_n e^{int}}{(in)^4}, \tag{7.13}$$

then

$$\int_{x_1,x_2,x_3,x_4}^{x} f(t)d_4t = H_4(F; x_1, x_2, x_3, x_4, x) \tag{7.14}$$

where $x_1 < x_2 < x_3 < x_4$, x are any five numbers.

THEOREM 5. *If*

$$\sum_{-\infty}^{\infty} c_n e^{inx},$$

has $c_n = o(n)$ *and is* $(C, 2)$ *summable to* $f(t)$ *then*

$$c_n = \frac{3}{8\pi^4}\int_{-4\pi,-2\pi,2\pi,4\pi}^{0} f(t)e^{-int}d_4t.$$

THEOREM 6. *Let* $f(x)$ *be the finite fourth Riemann symmetric derivative of a function continuous on* $[a, b]$ *and with a second derivative on* (a, b). *Let* $f(x)$ *be Lebesgue integrable except at a finite number of points* $\beta_1, \ldots, \beta_n$. *Further suppose that*

$$\lambda(x) = \int_{a_i}^{x}\int_{a_i}^{y}\int_{a_i}^{u} f(t)\,dt\,du\,dy \qquad i = 1, 2, \ldots, n$$

is Denjoy integrable in some interval (α_i, γ_i) *containing* β_i. *Then if we define*

$$\Phi(x) = \int_a^x\int_a^y\int_a^u\int_a^v f(t)\,dt\,dv\,du\,dy$$

then

$$\int_{x_1,x_2,x_3,x_4}^{x} f(t)d_4t = H_4(\Phi: x_1, x_2, x_3, x_4, x).$$

REFERENCES

1. R. D. James, *A generalized integral II*, Can. J. Math., *3* (1950), 297–306.
2. R. D. James, *Summable trigonometric series*, Pac. J. Math., *6* (1956), 99–110.
3. R. L. Jeffery, *Trigonometric series* (Toronto, 1956).
4. C. Kassimatis, *Functions which have generalized Riemann derivatives*, Can. J. Math., *3* (1958), 413–420.
5. S. Saks, *On the generalized derivative*, J. London Math. Soc., *7* (1932), 247–251.
6. S. Verblunsky, *The generalized third derivative and its application to the theory of trigonometric series*, Proc. London Math. Soc. (2), *31* (1930), 387–406.
7. S. Verblunsky, *The generalized fourth derivative*, J. London Math. Soc., *6* (1931), 82–84.
8. A. Zygmund, *Trigonometric series* (2nd ed.; Cambridge, 1959), I, II.

University of British Columbia

A NOTE ON NON-NEGATIVE MATRICES

C. R. PUTNAM

1. Introduction. This note can be regarded as an addendum to the paper **(4)**. On the complex Hilbert space of vectors $x = (x_1, x_2, \ldots,)$ a matrix A is said to be bounded if there exists a constant M such that $||Ax|| \leqslant M||x||$ whenever $||x||^2 = \sum |x_k|^2 < \infty$; the least such M is denoted by $||A||$. Only bounded matrices A and vectors x satisfying $||x|| < \infty$ will be considered in the sequel. The spectrum of A, denoted by $sp(A)$, is the set of values for which the resolvent $R(\lambda) = (A - \lambda I)^{-1}$ fails to be bounded. The notation $A \geqslant 0$ or $A > 0$, where $A = (a_{ij})$, means that, for all i and j, $a_{ij} \geqslant 0$ or $a_{ij} > 0$ respectively. There was stated in **(4)** the following theorem (also contained in some results of Bonsall, cf. the references cited in **(4)**) generalizing results of Perron and Frobenius for finite matrices:

(I) *If* $A \geqslant 0$, *then* $\mu = \sup |\lambda|$, *where* λ *is in* $\operatorname{sp}(A)$, *also belongs to* $\operatorname{sp}(A)$.

The proof in **(4)** of this theorem is not correct for arbitrary bounded $A \geqslant 0$, although it is valid for any such matrix with a spectrum identical with the set of (function theoretical) singularities of its resolvent, that is, with the set of singularities of at least one element of the resolvent. However, although the spectrum always contains this latter set, there exist bounded matrices, even satisfying $A \geqslant 0$, for which a number can belong to the spectrum and, at the same time, be an analytic point of each element of the resolvent. Such a matrix is given by $B = (b_{ij})$ where $b_{ij} = 1$ or 0 according as j is, or is not, $i + 1$. In fact, $B \geqslant 0$, $\operatorname{sp}(B)$ is the unit disk $|\lambda| \leqslant 1$, and 0 is the only singularity of $R(\lambda)$ (see **(6**, p. 145**)**). In view of this circumstance, an alternate simple proof of (I) will be given in § 2 below.

A few remarks relating to **(4)** will be made in § 3. In § 4, generalizations of certain theorems stated in a recent paper of Birkhoff and Varga **(1**, pp. 356-357**)**, will be given.

2. In order to prove (I), it is sufficient to show that $R(\lambda)$ is bounded whenever $R(|\lambda|)$ is bounded. Now $R(\lambda)$ is given by $R(\lambda) = -\sum A^n/\lambda^{n+1}$ whenever $|\lambda|$ is sufficiently large, in fact, whenever $|\lambda|$ exceeds the spectral radius,

$$\mu\left(= \lim_{n\to\infty} ||A^n||^{1/n}\right),$$

Received July 7, 1959. This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under Contract No. AF 18 (603) – 139. Reproduction in whole or in part is permitted for any purpose of the United States Government.

of A (cf., for example, **(5**, p. 421**)**). Since $A \geqslant 0$, also $A^n \geqslant 0$, and so if $R(\lambda) = (R_{ij}(\lambda))$, it is clear that, for $|\lambda| > \mu$, $|R_{ij}(\lambda)| \leqslant -R_{ij}(|\lambda|)$. If $x = (x_1, x_2, \ldots,)$ and $X = (|x_1|, |x_2|, \ldots,)$, then $||x|| = ||X||$ $(< \infty)$. Consequently, $||R(\lambda)x||^2 = \sum_i|\sum_j R_{ij}(\lambda)x_j|^2 \leqslant \sum_i(\sum_j R_{ij}(|\lambda|)|x_j|)^2 = ||R(|\lambda|)X||^2 \leqslant ||R(|\lambda|)||^2||x||^2$, and so $||R(\lambda)|| \leqslant ||R(|\lambda|)||$ whenever $|\lambda| > \mu$. But if μ is not in sp(A), then $||R(\mu)|| < \infty$, and it follows from the continuity of $||R(\lambda)||$ on the complement of the spectrum, and from the fact that $||R(\lambda)|| \to \infty$ whenever λ is not in sp(A) and tends to a point of sp(A), that $||R(\lambda)|| \leqslant ||R(\mu)|| < \infty$ for $|\lambda| = \mu$. Since sp(A) is closed, this last inequality implies that the spectral radius is less than μ, a contradiction, and the proof of (I) is now complete.

3. The third theorem in **(4)** can be stated as

(III) *If $A \geqslant 0$ and if μ of* (I) *is a pole of the resolvent* $R(\lambda) = (A - \lambda I)^{-1}$ *then there exists a characteristic vector $x \geqslant 0$ of A belonging to μ, thus $Ax = \mu x$* $(x \neq 0)$.

Actually it was assumed in **(4)** that μ should be positive; the proof given there, § 5, makes it clear, however, that this need not be assumed. If λ is real and satisfies $\lambda > \mu$, then $R(\lambda) = -\sum A^n/\lambda^{n+1} \leqslant 0$, and the matrix inequality $c_{-N} \leqslant 0$ $(N \geqslant 1,\ c_{-N} \neq 0)$ needed in the representation

$$R(\lambda) = \sum_{n=-N}^{\infty} c_n(\lambda - \mu)^n$$

of **(4**, § 5**)**, is still assured.

Whether the assumption in (III) that μ be a pole of $R(\lambda)$ can be weakened to the (implied) condition that μ be an isolated point of sp(A) and belong to the point spectrum will remain undecided. It is even conceivable that only the assumption that μ be in the point spectrum is needed in the hypothesis of (III).

Incidentally, the statement of **(3)**, and mentioned in **(4)**, that *if $A \geqslant 0$ and is completely continuous, and if the diagonal elements of every power A^n are zero, then zero is the only point of* sp(A), surely cannot be true if the assumption of complete continuity is omitted, as the matrix B cited earlier in this paper shows.

4. Generalizations of certain theorems for finite matrices stated in a recent paper of Birkhoff and Varga **(1**, pp. 256–257**)**, will be given in this section. Corresponding to the terminology of **(1)**, a matrix A will be called non-negative or positive according as $A \geqslant 0$ or $A > 0$, essentially non-negative if $a_{ij} \geqslant 0$ for $i \neq j$, and irreducible (also indecomposable, cf. the references cited in **(1)**) if, for any i and j, there exists a finite sequence $i = k(0), k(1), \ldots, k(N) = j$ for which $a_{k(n-1),k(n)} \neq 0$ for $k = 1, 2, \ldots, N$. A vector $x = (x_1, x_2, \ldots,)$ will be called non-negative or positive according as $x_i \geqslant 0$ or $x_i > 0$ for all i.

(i) *If A is essentially non-negative then* $\nu = \max Re(\mathrm{sp}(A))$ *is in* $\mathrm{sp}(A)$; *moreover,* $\nu > Re(\lambda)$ *if* $\lambda \neq \nu$ *and* λ *is in* $\mathrm{sp}(A)$. *In case* ν *is a pole of the resolvent* $R(\lambda) = (A - \lambda I)^{-1}$, *then A has a non-negative characteristic vector x belonging to* ν.

In fact, (i) follows readily from (I) and (III) if these latter theorems are applied to the matrix $C = A + \alpha I$ which is non-negative if α is positive and sufficiently large. It is to be noted that the resolvent of C is given by $R(\lambda - \alpha)$.
Furthermore,

(ii) *If A is essentially non-negative and irreducible and if ν of* (i) *is a pole of* $R(\lambda)$, *then* (a) *ν is a simple pole of* $R(\lambda)$, (b) *ν is a simple characteristic number, and* (c) *there exists a positive characteristic vector x of A belonging to* ν.

Assertion (ii) follows from (IV) of **(4)**, namely,

(IV) *If* $C \gg 0$, *if for every pair, i, j there exists an integer* $M = M(i,j)$ *such that* $(C^M)_{ij} > 0$, *and if μ of* (I) *is a pole of* $R(\lambda) = (C - \lambda I)^{-1}$, *then* (a) *$\mu$ is a simple pole of* $R(\lambda)$, (b) *μ is a simple characteristic number, and* (c) *there exists a characteristic vector* $x > 0$ *belonging to* μ.

In order to see this, let (IV) be applied to $C = A + \alpha I$, which is non-negative for α positive and sufficiently large, and note that the condition $(C^M)_{ij} > 0$ for some positive integer $M = M(i,j)$ is a consequence of the present assumption of irreducibility of A, provided that α is sufficiently large. (In this connection for finite matrices, see **(2**, p. 20**)**). For, let $\alpha > 0$ be chosen so large that the diagonal elements c_{ii} of C are positive. Since $c_{ij} = a_{ij}$ if $i \neq j$, it is then clear that the irreducibility of A implies that of C. Consequently, since $C \gg 0$, there exists for any pair i, j a positive integer $M = M(i,j)$ and a finite sequence $i = k(0), k(1), \ldots, k(M) = j$ such that

$$d = \prod_{n=1}^{M} c_{k(n-1),k(n)} > 0.$$

But $(C^M)_{ij}$ is given by a sum of non-negative terms one of which is d and so $(C^M)_{ij} > 0$. Thus, as remarked above, (ii) follows from (IV) of **(4)**. Incidentally, the above argument makes clear that a non-negative matrix, here C, satisfies $(C^M)_{ij} > 0$ for every pair i, j and some positive integer $M = M(i,j)$ if and, in fact, only if, it is irreducible.

References

1. G. Birkhoff and R. S. Varga, *Reactor criticality and nonnegative matrices*, J. Soc. Indust. Appl. Math., *6* (1958), 354–377.
2. I. N. Herstein, *A note on primitive matrices*, Amer. Math. Monthly, *61* (1954), 18–20.
3. M. G. Krein and M. A. Rutman, *Linear operators leaving invariant a cone in a Banach space*, Uspehi Matem., Nauk (N.S.), *3* (23) 1948), 3–95; Amer. Math. Soc. Translation No. 26 (page reference in paper refers to this translation).
4. C. R. Putnam, *On bounded matrices with non-negative elements*, Can. J. Math., *10* (1958), 587–591.
5. F. Riesz and B. Sz-Nagy, *Leçons d'analyse fonctionnelle* (Budapest, 1953).
6. A. Winter, *Spektraltheorie der unendlichen Matrizen* (Leipzig, 1929).

Purdue University

SUMMABILITY METHODS ON MATRIX SPACES

JOSEPHINE MITCHELL

§1. Introduction. The matrix spaces under consideration are the four main types of irreducible bounded symmetric domains given by Cartan **(5)**. Let $z = (z_{jk})$ be a matrix of complex numbers, z' its transpose, z^* its conjugate transpose and $I = I^{(n)}$ the identity matrix of order n. Then the first three types are defined by

$$D = [z | I - zz^* > 0], \tag{1}$$

where z is an n by m matrix ($n < m$), a symmetric or a skew-symmetric matrix of order n **(16)**. The fourth type is the set of complex spheres satisfying

$$|z'z| < 1, 1 - 2z^*z + |z'z|^2 > 0,$$

where z is an n by 1 matrix. It is known that each of these domains possesses a distinguished boundary B which in the first three cases is given by

$$B = [u | uu^* = I]. \tag{2}$$

(In the case of skew symmetric matrices the distinguished boundary is given by (2) only if n is even.)

In § 2 we consider the following problem for the first type of domain with $m = n$, in which case u is a unitary matrix, the (real) dimension of B is n^2 and of D is $2n^2$. Let $f(u)$ be a real integrable function defined on B and consider the integral operator

$$I(f, z) = \int_B P(z, u) f(u)\, dV, \tag{3}$$

where $P(z, u)$ is the Poisson kernel **(14)**

$$P(z, u) = V^{-1} \det{}^n (I - zu^*)^{-1} (I - zz^*)(I - uz^*)^{-1}, \tag{4}$$

V is the Euclidean volume of B, and dV the Euclidean volume element. It is known that $I(f, z)$ is a harmonic function of z if $I - zz^* > 0$ or $I - zz^* < 0$. (I proved this fact in **(14)** for $z \in D$ but the proof is valid for all z and $u \in B$ for which $\det(I - zu^*) \neq 0$. It is easily proved that $\det(I - zu^*) \neq 0$ for $u \in B$ and all z such that $I - zz^* > 0$ or $I - zz^* < 0$.) Here a harmonic function is a function of class C^2, which satisfies on D the Laplace equation corresponding to the invariant metric of D, that is, the metric invariant with respect to the group of 1 to 1 analytic transformations mapping D onto itself **(14)**. This invariant metric is given by

$$ds^2 = \sigma[(I - zz^*)^{-1} dz (I - z^*z)^{-1} dz^*],$$

Received November 30, 1959. Presented to meetings of American Mathematical Society on January 29, 1960 and January 24, 1961.

where $\sigma(A)$ is the trace of the matrix A and $dz = (dz_{jk})$, and the corresponding Laplace equation is

$$4\sigma[\bar{\partial}(I - z^*z)\partial'(I - zz^*)] = 0, \qquad \partial = (\partial/\partial z_{jk}).$$

It has been proved by Hua and Lowdenslager that given a real function f, continuous on B, there exists a function F, harmonic on D, such that $F(z) \to f(u_0)$ as $z \to u_0 \in B$ radially, that is, along the set ρu_0, $0 < \rho < 1$ **(6; 9)**. Further, if F is continuous on the closure $\bar{D}$ of D and satisfies certain other conditions due to Lowdenslager on the boundary of D other than B, then F is unique **(8)**. Now for the particular case of the unit circle, $z\bar{z} = 1$, if we merely assume that f is integrable on it, then $I(f, \rho u_0) \to f(u_0)$ as $\rho \to 1 (0 < \rho < 1)$ **(20)**; this method of approach is known as *Abel–Poisson summability of Fourier series*. We prove this result for matrix spaces. (See note added in proof).

In § 3 we consider for the first type of domain ($n \leqslant m$) some properties of complete orthonormal systems (CONS) of complex homogeneous polynomials defined on D. The space D is circular with center at $z = 0$, that is, if $z \in D$, then $e^{i\theta}z \in D$ for $0 \leqslant \theta < 2\pi$. Hence any two powers

$$P(z) = \prod_{j,k} z_{jk}^{s_{jk}}, \qquad Q(z) = \prod_{j,k} z_{jk}^{t_{jk}}, \tag{5}$$

s_{jk}, t_{jk} non-negative integers, for which

$$\sum_{j,k} s_{jk} \neq \sum_{j,k} t_{jk}, \tag{6}$$

are orthogonal, that is,

$$(P, Q) = \int_D P(z)\bar{Q}(z)dW = 0 \tag{7}$$

(dW is Euclidean volume element on D) **(6)**. Also if $f \in$ class L^2 on D, which means that f is single-valued and analytic on D and has finite norm $||f|| = [(f, f)]^{\frac{1}{2}}$ **(2)**, then the set $\{P\}$ is complete with respect to functions of class L^2 **(6)**.

Here we refine conditions (6) to show that $(P, Q) \neq 0$ implies

$$\sum_k s_{\nu k} = \sum_k t_{\nu k}, \qquad \sum_j s_{j\mu} = \sum_j t_{j\mu} \qquad (\nu = 1, \ldots, n; \mu = 1, \ldots, m). \tag{8}$$

By means of (8) the set of powers $\{P\}$ is subdivided into disjoint subsets whose members need not be orthogonal to each other. The elements of a subset are made into an orthonormal set by the Gram Schmidt formulas, thus giving a CONS of homogeneous polynomials $\{\phi\}$ on D. We note that Hua has constructed a CONS of functions of class L^2 on D using representation theory **(6)**.

In § 4 applications of the CONS $\{\phi_\nu\}$ are given. First an Abelian theorem is obtained and then a Tauberian theorem for the orthogonal series $\sum a_\nu\phi_\nu$ as z approaches $[I, 0]$ of B along the matrix $[r, 0]$ where r is the diagonal matrix $[r_1, \ldots, r_n]$, $0 \leqslant r_j < 1$. Next a Cauchy's inequality is obtained for the Fourier coefficients a_ν. Finally two mean value theorems, which generalize analogous theorems for the unit circle, are proved.

§2. Poisson summability.

1. *Reduction of integral* (1.3) *to normal form*. Rauch outlined this reduction to me. The transformation

$$w = zu_0^{-1}, \qquad u_0u_0^* = I, \tag{1}$$

takes $z = u_0$ into $w = I$ and also leaves D and B invariant since under it $I - ww^* = I - zz^*$. Also if $u \to v$ under (1) $P(z, u) \to P(w, v)$ and

$$dV_u = \frac{\dot{u}}{\det^n u} = \frac{\partial(v)}{\partial(u)} \frac{\dot{v}}{\det^n v \det^n u_0},$$

where

$$\dot{u} = (-1)^{-\frac{1}{2}n(n+1)} \prod_{j,k} du_{jk}$$

(14). Now $du = dv\, u_0$, the Jacobian of which is $\partial(u)/\partial(v) = \det^n u_0$ **(3)**. Thus $dV_u \to dV_v$. Also $f(u) \to f(vu_0) = f_0(v)$ so that $I(f, z) \to I(f_0, w)$.

If $w \to I$ along the set of points ρI $(0 < \rho < 1)$, then

$$\det(I - ww^*) = (1 - \rho^2)^n$$

and

$$\begin{aligned} Q = (I - wv^*)(I - vw^*) &= I - wv^* - vw^* + ww^* \\ &= I(1 + \rho^2) - \rho(v + v^*). \end{aligned}$$

Now v is unitary equivalent to a diagonal matrix v_D which is also unitary **(10**, Theorem 41.41**)**, that is,

$$v = U^* v_D U. \tag{2}$$

Thus if $v_D = [d_1, \ldots, d_n]$, then $d_j\bar{d}_j = 1$ and we can write $d_j = e^{i\theta_j}(0 < \theta_j < 2\pi)$. Hence

$$Q = U^*[I(1 + \rho^2) - \rho(v_D^* + v_D)]U$$

and

$$\det Q = \det[I(1 + \rho^2) - \rho(v_D^* + v_D)] = \prod_{j=1}^{n} (1 - 2\rho\cos\theta_j + \rho^2).$$

Let

$$p(\rho, \theta_j) = \frac{1 - \rho^2}{1 - 2\rho\cos\theta_j + \rho^2}, \tag{3}$$

which is 2π times the Poisson kernel for the unit circle. Then

$$P(\rho I, v) = V^{-1} \prod_{j=1}^{n} p^n(\rho, \theta_j) \tag{4}$$

and

$$I(f_0, \rho I) = V^{-1} \int_B \prod_j p^n(\rho, \theta_j) f_0(v)\, dV_v. \tag{5}$$

According to Weyl **(19**, p. 197**)** if (2) holds, then

$$dV_v = [dV_U]\Delta\bar{\Delta}\, d\theta_1 \ldots d\theta_n, \tag{6}$$

where

$$\Delta\bar{\Delta} = \prod_{j<k} |e^{i\theta_j} - e^{i\theta_k}|^2 = 4 \prod_{j<k} \sin^2\tfrac{1}{2}(\theta_j - \theta_k) \tag{7}$$

and $dV_0 = [dV_V]$ is a constant times the Euclidean volume element on the other $n^2 - n$ parameters defining B. Let B_0 be this part of B. Now $P(\rho I, v)$, considered as a function of v, is independent of the other $n^2 - n$ parameters and hence

$$I(f_0, \rho I) = \int_0^{2\pi} \cdots \int_0^{2\pi} P(\rho I, v) F(\theta) \Delta\bar{\Delta} d\theta_1 \ldots d\theta_n, \tag{8}$$

where

$$F(\theta) = F(\theta_1, \ldots, \theta_n) = \int_{B_0} f_0(v) dV_0. \tag{9}$$

2. *Convergence theorem for* (8). It is sufficient to consider (8) for $n = 2$, in which case replacing θ_1 by s and θ_2 by t

$$I(f_0, \rho I) = 4V^{-1} \int_0^{2\pi} \int_0^{2\pi} p^2(\rho, s) p^2(\rho, t) \sin^2\tfrac{1}{2}(s - t) F(s, t) ds dt \tag{10}$$

and we consider $\lim_{\rho\to 1} I(f_0, \rho I)$. Subtracting S from each side we reduce (10) by well-known methods in Fourier series **(20)** to a consideration of

$$I = 4V^{-1} \int_0^{\pi} \int_0^{\pi} p^2(\rho, s) p^2(\rho, t) \sin^2\tfrac{1}{2}(s - t) G(s, t) ds dt \tag{11}$$

and

$$I_1 = 4V^{-1} \int_0^{\pi} \int_0^{\pi} p^2(\rho, s) p^2(\rho, t) \sin^2\tfrac{1}{2}(s + t) G(2\pi - s, t) ds dt, \tag{11a}$$

where

$$G(s, t) = F(s, t) + F(2\pi - s, 2\pi - t) - 2V_0 S, \tag{12}$$

V_0 being the volume of B_0. We show that under certain hypotheses on $G(s, t)$, $I \to 0$ as $\rho \to 1$ and the proof is similar for I_1. In (11) integrate by parts with respect to s and t. Since

$$H(s, t) = \iint G(s, t) ds dt$$

is zero for $s = 0$ or $t = 0$, $\sin^2\frac{1}{2}(s - t) = 0$ for $s = t = \pi$, $\sin^2\frac{1}{2}(\pi - \theta) = \cos^2\frac{1}{2}\theta$ and $p(\rho, \pi) = (1 - \rho)/(1 + \rho)$, we get

$$\begin{aligned} I = &- \left(\frac{1 - \rho}{1 + \rho}\right)^2 \int_0^{\pi} \frac{\partial}{\partial t} [p^2(\rho, t) \cos^2\tfrac{1}{2}t] H(\pi, t) dt \\ &- \left(\frac{1 - \rho}{1 + \rho}\right)^2 \int_0^{\pi} \frac{\partial}{\partial s} [p^2(\rho, s) \cos^2\tfrac{1}{2}s] H(s, \pi) ds \\ &+ \int_0^{\pi} \int_0^{\pi} \frac{\partial^2}{\partial s \partial t} [p^2(\rho, s) p^2(\rho, t) \sin^2\tfrac{1}{2}(s - t)] H(s, t) ds dt \\ \equiv &\; I_1 + I_2 + I_3. \end{aligned} \tag{13}$$

The following inequalities are used in considering I_1, I_2, and I_3 **(20)**:

(14) (i) $0 < p(\rho, \theta) < \dfrac{1+\rho}{1-\rho}$

(ii) $0 < p(\rho, \theta) < \dfrac{1-\rho^2}{4\sin^2\frac{1}{2}\theta}$

(iii) $\lim_{\theta\to 0} \dfrac{\theta}{\sin\theta} = 1$

(iv) $\int_0^\pi p(\rho, \theta)d\theta = \pi$

(v) $\int_\delta^\pi p(\rho, \theta)d\theta = o(1)$ as $\rho \to 1$ for $0 < \delta < \pi$.

Also

$$\frac{\partial}{\partial\theta} p(\rho, \theta) = 2\rho(1-\rho^2)\sin\theta\, d^{-2}(\rho, \theta)$$

where

$$d(\rho, \theta) = 1 - 2\rho\cos\theta + \rho^2 = \frac{1-\rho^2}{p(\rho, \theta)}$$

and

$$\begin{aligned}\frac{\partial^2}{\partial s\partial t}&[p^2(\rho, s)p^2(\rho, t)\sin^2\tfrac{1}{2}(s-t)] = p(\rho, s)p(\rho, t)\,.\\ &[16\rho^2(1-\rho^2)^2\sin s\sin t\sin^2\tfrac{1}{2}(s-t)d^{-2}(\rho, s)d^{-2}(\rho, t) - \tfrac{1}{2}p(\rho, s)p(\rho, t)\\ &.\cos(s-t) - 2\rho(1-\rho^2)\sin s\sin(s-t)p(\rho, t)d^{-2}(\rho, s)\\ &+ 2\rho(1-\rho^2)\sin t\sin(s-t)p(\rho, s)d^{-2}(\rho, t)].\end{aligned}$$

Concerning the function $G(s, t)$ we assume that

(15) $$\lim_{h,k\to 0} \frac{1}{hk}\int_0^h\int_0^k |G(s, t)|dsdt = 0$$

$$\int_0^h\int_0^k |G(s, t)||dsdt| < C|hk|, \qquad (0 < |h|, |k| < \pi),$$

C an absolute constant. (See **(11)** where these conditions were used in a similar connection.) Then using (15) we find

$$|I_1| < \pi(1-\rho^2)C\int_0^\pi |2p(\rho, t)\cos^2\tfrac{1}{2}t\,\partial p(\rho, t)/\partial t - p^2(\rho, t)\sin\tfrac{1}{2}t\cos\tfrac{1}{2}t|tdt$$

$$= 0((1-\rho)\int_0^\pi p(\rho, t)dt) = 0(1-\rho)$$

and similarly for I_2. Now

$$I_3 = \int_0^\delta\int_0^\delta + \left\{\int_0^\delta\int_\delta^\pi + \int_\delta^\pi\int_0^\delta + \int_\delta^\pi\int_\delta^\pi\right\} = I_{31} + I_{32}.$$

By (15) it is found that

$$|I_{31}| < \epsilon \int_0^{\pi}\int_0^{\pi} p(\rho, s)p(\rho, t)dsdt,$$

where ϵ is an arbitrary positive number, and

$$I_{32} = 0\left(\frac{1-\rho}{\sin^2\frac{1}{2}\delta}\right).$$

Consequently given $\epsilon > 0$ choose $\delta > 0$ so that

$$\left|\int_0^{s}\int_0^{t} G(s, t)dsdt\right| < \epsilon|st|,$$

if $|s| < \delta$ and $|t| < \delta$. With fixed δ choose $1 - \rho$ sufficiently small. Then $I = 0(\epsilon)$ for ρ sufficiently close to 1. Consequently we have proved

THEOREM 2.1. *Let $F(s, t)$ be an integrable function. If $G(s, t) = F(s, t) + F(2\pi - s, 2\pi - t) - 2V_0S$ satisfies conditions* (15), *then*

$$\lim_{\rho\to1^-} 4V^{-1}\int_0^{2\pi}\int_0^{2\pi} p^2(\rho, s)p^2(\rho, t)\sin^2\tfrac{1}{2}(s - t)F(s, t)dsdt = S.$$

For $n > 2$ it would be sufficient to assume that

$$\text{(16)}\quad \lim_{\theta_j\to0}\frac{1}{\theta_1\ldots\theta_n}\int_0^{\theta_n}\ldots\int_0^{\theta_n}|G(\theta)|d\theta_1\ldots d\theta_n = 0,$$

$$\int_0^{\theta_1}\ldots\int_0^{\theta_n}|G(\theta)||d\theta_1\ldots d\theta_n| < C|\theta_1\ldots\theta_n|, \; (0 < |\theta_j| < \pi, j = 1, \ldots, n),$$

where $G(\theta)$ is defined similarly to $G(s, t)$. We obtain

THEOREM 2.2. *Let f be an integrable function on the unitary group B such that the function $F(\theta)$ defined by* (9), *where f_0 is the transform of f under* (1), *satisfies* (16). *Then the Poisson integral* (1.3) *has a limit if z approaches the point u_0 on B radially.*

§3. Complete orthonormal systems on D. Orthogonal developments.

1. *Integration over D.* Let z be an n by m matrix ($n < m$), z_p its pth row and Z_p the submatrix consisting of the first p rows ($p = 1, \ldots, n$). The inner product (f, g) defined by (1.7) may be transformed into an iterated integral over the product of n hyperspheres by a procedure due to Hua **(12)**, giving

$$\text{(1)}\quad (f, g) = \left(\frac{1}{2i}\right)^{mn}\int_{w_1w_1^*<1}(1 - w_1w_1^*)^{n-1}\dot{w}_1\ldots\int_{w_nw_n^*<1} f\bar{g}\,\dot{w}_n,$$

where

$$\text{(2)}\quad \dot{w}_p = \prod_{k=1}^{m} dw_{pk}d\bar{w}_{pk}$$

$$w_p = z_p\Gamma_{p-1}$$

$$z_p = w_p\Gamma_{p-1}^{-1} \qquad (p = 1, \ldots, n, \Gamma_0 = 1),$$

and Γ_{p-1} is a unique positive definite matrix such that

$$\Gamma_{p-1}\Gamma_{p-1}^{*} = (I - Z_{p-1}^{*}Z_{p-1})^{-1}. \tag{3}$$

2. *Construction of the matrix* $\Gamma_{p-1}{}^{-1}(2 < p < n)$. Let $U_p(q) = U(q)$ be the minor formed from the first q rows and columns of

$$U_p = U = (u_{jk}) = I - Z_{p-1}^{*}Z_{p-1}, \tag{4}$$

$\mathscr{U}_p(q) = \mathscr{U}(q)$ the corresponding submatrix and $u_{m-j} = (u_{m-j,1}, \ldots, u_{m-j,m-j-1})$. Since $I - Z_{p-1}Z^{*}{}_{p-1}$ is the leading $(p-1)$th principal submatrix of $I - zz^{*}$, it is positive definite, hence U is positive definite and $U(q)$ are positive. Thus the hermitian matrix U may be reduced to diagonal form by the well-known Kronecker reduction **(1)** whose $(j+1)$th step is

$$V_{j+1} \ldots V_1 U V_1^{*} \ldots V_{j+1}^{*} = \begin{pmatrix} \mathscr{U}(m-j-1) & & 0 & \\ & X_{m,m-j} & & \\ & & \ddots & \\ 0 & & & X_{mm} \end{pmatrix}.$$

$(0 < j < m-2)$, where

$$V_{j+1} = \begin{pmatrix} I^{(m-j-1)} & 0 \\ -u_{m-j}\mathscr{U}^{-1}(m-j-1) & I^{(j+1)} \\ 0 & \end{pmatrix},$$

$$X_{m,m-j} = U(m-j)U^{-1}(m-j-1),$$

$$X_{m1} = U(1).$$

Also

$$u_{m-j}\mathscr{U}^{-1}(m-j-1) = U^{-1}(m-j-1)\cdot$$

$$\left((-1)^{k+m-j+1}\,U\begin{pmatrix}1, \ldots, [k], \ldots, m-j \\ 1, \ldots, m-j-1\end{pmatrix}\right), \quad (1 < k < m-j-1),$$

where

$$U\begin{pmatrix}1, \ldots, [k], \ldots, m-j \\ 1, \ldots, m-j-1\end{pmatrix}$$

is the minor of U formed from rows $1, \ldots, m-j$ with row k omitted and columns $1, \ldots, m-j-1$.

Now $V_{j+1}{}^{-1}$ equals V_{j+1} with the sign of the matrix $-u_{m-j}\mathscr{U}^{-1}(m-j-1)$ changed. Also X_{mk} is real. Hence we can take

$$\Gamma_{p-1}^{*-1} = V_1^{-1} \ldots V_{m-1}^{-1}[X_{m1}^{\frac{1}{2}}, \ldots, X_{mm}^{\frac{1}{2}}]. \tag{5}$$

By an inductive proof we may show that

$$(6)\quad V_1^{-1}\dots V_{m-1}^{-1} = \left(\frac{U\binom{j}{1}}{U(1)}\,\frac{U\binom{1,j}{1,2}}{U(2)}\cdots\frac{U\binom{1,2,\dots,[j-1],j}{1,2,\dots,j-1}}{U(j-1)}\,1\,0\dots 0\right),$$

$(1 < j < m)$. (For details of the proof cf. **(15)**.) From (4) it follows that

$$(7)\quad U\binom{1,\dots,[i],r}{1,\dots,i} = U_p\binom{1,\dots,[i],r}{1,\dots,i}$$

$$= \det\left(\delta_{jk} - \sum_{l=1}^{p-1} \bar{z}_{lj}z_{lk}\right),$$

$$(j = 1,\dots,i-1,r;\, k = 1,\dots,i;\, i+1 < r < m,\, 1 < i < m-1).$$

3. *Formula for z_{pr}.* From (2), (5), and (6) follows

$$(8)\quad z_{pr} = \sum_{i=1}^{r-1} c_{pi}\bar{U}_p\binom{1,\dots,[i],r}{1,\dots,i}w_{pi} + c_{prr}w_{pr} \quad (p > 2)$$

$$z_{1r} = w_{1r}\left(1 < r < m;\ \sum_{i=1}^{0} = 0\right),$$

where

$$(9)\quad \begin{aligned} c_{pi} &= [U_p(i)U_p(i-1)]^{-\frac{1}{2}} \quad (1 < i < r-1)\\ c_{prr} &= [U_p(r)/U_p(r-1)]^{\frac{1}{2}}\\ c_{1i} &= c_{1rr} = 1. \end{aligned}$$

The formula

$$(10)\quad U_p(q) = \prod_{k=1}^{p-1}\left(1 - \sum_{j=1}^{q} w_{kj}\bar{w}_{kj}\right) \qquad (p > 2)$$

holds.

We prove (10) by induction on p. Since by (8)

$$U_2(q) = \det(I - z_1^* z_1) = \det(I - z_1 z_1^*) = 1 - \sum_{j=1}^{q} w_{1j}\bar{w}_{1j},$$

where $z_1 = (z_{11},\dots, z_{1q})$, (10) is true for $p = 2$. Assume (10) holds for $U_p(q)$ and prove for $U_{p+1}(q)$. Since $U_p(q) \neq 0$, there exists a unimodular matrix A such that the matrix $\mathscr{V}_{p+1}(q) = (\delta_{jk} - \sum_{l=1}^{q} z_{jl}\bar{z}_{kl})$ $(1 < j, k < p)$ is equal to

$$A[\mathscr{V}_p(q),\, 1 - z_p\mathscr{U}_p^{-1}(q)z_p^*]A^*,$$

(12), and thus

$$U_{p+1}(q) = \det \mathscr{V}_{p+1}(q) = \det \mathscr{V}_p(q)(1 - z_p\mathscr{U}_p^{1}(q)z_p^*).$$

Since $\det \mathscr{V}_p(q) = U_p(q)$, using (10) we need only consider the last factor on the right, which equals

$$E = 1 - \sum_{j,k=1}^{q} z_{pj}U_{kj}\bar{z}_{pk}/U_p(q),$$

where U_{kj} is the cofactor of the element u_{kj} in the matrix $\mathscr{U}_p(q)$. Substitute (8) for z_{pj} and $\bar{z}_{pk}$. The term w_{pr} occurs when $j = r$ or when $i = r$ and $j = r + 1, \ldots, q$ and $\bar{w}_{p\lambda}$ occurs when $k = \lambda$ or $i = \lambda$ and $k = \lambda + 1, \ldots, q$. Thus the coefficient, $C_{r\lambda}$, of $w_{pr}\bar{w}_{p\lambda}(1 \leqslant r, \lambda \leqslant q)$ is

$$C_{r\lambda} = U_p^{-1}(q)\left[c_{pr}\sum_{j=r+1}^{q} \bar{U}_p\begin{pmatrix}1, \ldots, [r], j\\ 1, \ldots, r\end{pmatrix}D_{j\lambda} + c_{prr}D_{r\lambda}\right],$$

where

$$D_{j\lambda} = c_{p\lambda}\sum_{k=\lambda+1}^{q} U_{kj}U_p\begin{pmatrix}1, \ldots, [\lambda], k\\ 1, \ldots, \lambda\end{pmatrix} + U_{\lambda j}c_{p\lambda\lambda}.$$

By means of elementary properties of determinants it is not difficult to prove in case $\lambda \leqslant r$ that $D_{j\lambda} = 0$ for $j = r$, $\lambda < r$ and $j = r + 1, \ldots, q$, $\lambda \leqslant r$. Hence $C_{r\lambda} = 0$ for $\lambda \neq r$. Also $C_{rr} = 1$. A similar proof holds for $r < \lambda$. Thus

$$E = 1 - \sum_{j=1}^{q} w_{pj}\bar{w}_{pj}$$

and $U_{p+1}(q)$ has the desired form, which proves (10).

4. *Structure of the* CONS *on D.*

THEOREM 3.1. $(P, Q) \neq 0$ *implies equations* (1.8).

Proof. We first show that

$$z_{pr} = w_{pr}\sum B_t w_{pt}\bar{w}_{pr}\prod (w_{\lambda_1\alpha_1}\bar{w}_{\lambda_1\beta_1}\cdots w_{\lambda_l\alpha_l}\bar{w}_{\lambda_l\beta_l}w_{\lambda_t r}\bar{w}_{\lambda_t\beta_t}), \tag{11}$$

where B_t is a function of $w_{jk}\bar{w}_{jk}$ $(1 \leqslant j \leqslant p - 1)$; $\lambda_1, \ldots, \lambda_l, \lambda_t$ take on values in the set $1, 2, \ldots, p - 1$; $\alpha_1, \ldots, \alpha_l$ is a subset of $1, \ldots, i - 1$ and $(\beta_1, \ldots, \beta_l, \beta_t)$ is a permutation of $(\alpha_1, \ldots, \alpha_l, i)$ $(i = 1, \ldots, r - 1)$. (Notice that each term of (11) can be grouped into pairs $w_{\alpha\beta}\bar{w}_{\gamma\delta}$ in two ways: (i) each pair belongs to the same row, (ii) each pair belongs to the same column.) Since $z_{1r} = w_{1r}$, (11) holds for $p = 1$. Now assume (11) for $p - 1$, $1 \leqslant r \leqslant m$, and prove for p. Upon expanding

$$\bar{U}_p\begin{pmatrix}1, \ldots, [i], r\\ 1, \ldots, i\end{pmatrix}$$

(given by (7)) and multiplying out the resulting factors, we find that its general term is

$$z_{\lambda_1\alpha_1}\bar{z}_{\lambda_1\beta_1}\cdots z_{\lambda_s\alpha_s}\bar{z}_{\lambda_s\beta_{\alpha_s}}z_{\lambda_{s+1}r}\bar{z}_{\lambda_{s+1}\beta_i},$$

where $\lambda_\alpha(\alpha = 1, \ldots, s + 1)$ takes on values from $1, 2, \ldots, p - 1$; $\alpha_1, \ldots, \alpha_s$ is a subset of $1, \ldots, i - 1$ and

$$\beta_{\alpha_1}, \ldots, \beta_{\alpha_s}, \beta_i'$$

is a permutation of $\alpha_1, \ldots, \alpha_s, i$. Thus the general term of z_{pr} would be w_{pr} times

$$c'_{pi} z_{\lambda_1 \alpha_1} \bar{\bar{z}}_{\lambda_1 \beta \alpha_1} \cdots z_{\lambda_s \alpha_s} \bar{\bar{z}}_{\lambda_s \beta \alpha_s} z_{\lambda_s+1\, r} \bar{\bar{z}}_{\lambda_s+1\, \beta_i} w_{pi} \bar{w}_{pr},$$

$c_{pi}' = c_{pi}/w_{pr}\bar{w}_{pr}$. Replace

$$z_{\lambda_1 \alpha_1 \lambda} \bar{\bar{z}}_{\lambda_1 \beta \alpha_1} \cdots \text{ by } w_{\lambda_1 \alpha_1} \bar{w}_{\lambda_1 \beta \alpha_1} \cdots$$

times a factor which by induction already has the required form and (11) follows.

Now consider $z_{pr}^{s_{pr}}$. From (11) we see that except for the first factor $w_{pr}^{s_{pr}}$ if $w_{\nu j}^{\sigma}$ occurs, then a factor $\bar{w}_{\nu k}^{\sigma}$ also occurs and if $w_{i\mu}^{\tau}$ appears, then $\bar{w}_{i\mu}^{\tau}$ also appears. Consequently in the expression for $z_{pr}^{s_{pr}}$ for each $\nu (\nu = 1, \ldots, p-1)$ the sum of the exponents of the factors $w_{\nu j} (j = 1, \ldots, m)$ equals the sum of the exponents of the $\bar{w}_{\nu k}$ $(k = 1, \ldots, m)$ and the sum of the exponents of w_{pj} equals the sum of the exponents of $\bar{w}_{pk}$ increased by s_{pr}. Similarly for the columns. Thus P can be expressed in the form

$$P(z) = P_0(w\bar{w}) \prod w_{jk}^{s_{jk}} = P_0(w\bar{w})P(w), \tag{12}$$

where $P_0(w\bar{w})$ contributes the same exponents to the sum of the elements in the νth row of w and of $\bar{w}$ and similarly for the columns of w and $\bar{w}$. An analogous expression holds for Q.

In (P, Q) replace P and Q by (12). Since $\{w_{\nu k}^{s_{\nu k}}\}$ $(k = 1, \ldots, m)$ forms an orthogonal set on $w, w^*, < 1$ **(2)**, $(P, Q) \neq 0$ if and only if for each k the exponent of $w_{\nu k}$ equals the exponent of $\bar{w}_{\nu k}$. Consequently if we sum the exponents of the νth row, owing to the form of $P_0(w\bar{w})$, $Q_0(w\bar{w})$ we obtain the first of equations (1.8) and summing the exponents of the μth column the second of equations (1.8). Thus the theorem is proved.

5. CONS. *Orthogonal development.* A CONS is constructed from the set of powers $\{P(z)\}$ as follows. Let $\alpha = (\alpha_1, \ldots, \alpha_{m+n})$ be a set of non-negative integers with $\sum_{j=1}^{n} \alpha_j = \sum_{k=1}^{m} \alpha_{k+n} = p$. The powers of the set $S(\alpha) = S(\alpha_1, \ldots, \alpha_{m+n})$ such that

$$\sum_k s_{\nu k} = \alpha_\nu, \qquad \sum_j s_{j\mu} = \alpha_{\mu+n}$$

need not be orthogonal to each other. (There exist sets $S(\alpha)$ whose members are not all orthogonal to each other—for example, in the 2 by 2 case the elements $z_{11}z_{22}$, $z_{12}z_{21}$ are not orthogonal **(12)**.) However if $P \in S(\alpha)$, $Q \in S(\beta)$, where $\alpha_j \neq \beta_j$ for some j, then by Theorem 3.1 $(P, Q) = 0$. We order the elements of the set $S(\alpha)$ in some convenient manner into a sequence $P_0, P_1, \ldots, P_{p(\alpha)}$. An ONS is constructed from these elements by the Gram-Schmidt formulas

$$\phi_\nu^{(p)}(z) = \det \begin{pmatrix} P_0 \ldots P_\nu \\ a_{00} \ldots a_{\nu 0} \\ \ldots \\ a_{0,\nu-1} \ldots a_{\nu,\nu-1} \end{pmatrix} / (D_{\nu-1}D_\nu)^{\frac{1}{2}}, \tag{13}$$

where

$$D_\mu = \det(a_{\alpha\beta}) \ (0 \leq \alpha, \beta \leq \mu;\ \mu = \nu - 1 \text{ or } \nu,\ \nu \neq 0),\ D_{-1} = 1.$$
$$a_{ij} = (P_i, P_j),\ (0 \leq i, j \leq \nu).$$

Now order the system $\{\phi_\nu^{(p)}\}$ into a sequence $\phi_1, \phi_2, \ldots$. The orthogonal development of any $f \in L^2$ with respect to the ONS is

$$\sum a_q \phi_q, \tag{14}$$

where a_q are the Fourier coefficients, (f, ϕ_q), of f. From Bergman's theory **(2)** it is known that (14) converges absolutely and continuously to f on D (continuous convergence means that the series converges uniformly on any compact set contained in D).

§4. Applications of the CONS.

1. *Abelian and Tauberian theorems.* Let $\{a_q\}$ be an arbitrary sequence of numbers and consider the behaviour of (3.14) as $z \in D$ approaches a point $u_0 \in B$. In particular let $u_0 = [I, 0]$, $I = I^{(n)}$, and $z \to u_0$ along the set of points $[r, 0]$ where r is the diagonal matrix $[r_1, \ldots, r_n]$, $0 \leq r_j < 1$. When $z = [r, 0]$ it is seen that P_ν of the set $S(\alpha)$ is either equal to 0 or to

$$\prod r_j^{p_j}$$

in case $\alpha_j = \alpha_{j+n} = p_j$ for $j = 1, \ldots, n$ and $\alpha_{j+n} = 0$ for $j > n$. Consequently for a fixed set α either all P_ν are zero or there is one P_ν different from zero in case $\alpha_j = \alpha_{j+n}$ for $j = 1, \ldots, n$ and $\alpha_{j+n} = 0$ for $j > n$. Order the elements of the set $S(\alpha)$ so that the non-zero term is the last term of the set. Then in (3.13) only $\phi_t^{(p)}(z)$, $t = p(\alpha)$, is different from zero when $z = [r, 0]$ and

$$\phi_t^{(p)}(z) = [D_{t-1}/D_t]^{\frac{1}{2}} P_t(z) = [D_{t-1}/D_t]^{\frac{1}{2}} r_1^{p_1} \ldots r_n^{p_n}.$$

Thus (3.14) reduces to a multiple power series. Let this series be summed by the usual method for power series. Then

$$S(r) = \sum_{p_1=0}^{\infty} \cdots \sum_{p_n=0}^{\infty} c_{p_1 \ldots p_n} r_1^{p_1} \ldots r_n^{p_n}, \tag{1}$$
$$c_{p_1 \ldots p_n} = a_q / (D_{t-1}/D_t)^{\frac{1}{2}},$$

where $\phi_t^{(p)} = \phi_q$ is the ordering of the ONS $\{\phi_t^{(p)}\}$ into a simple sequence. Let

$$S_{q_1 \ldots q_n}$$

be the partial sum of $S(r)$:

$$S_{q_1 \ldots q_n} = \sum_{p_1=0}^{q_1} \cdots \sum_{p_n=0}^{q_n} c_{p_1 \ldots p_n} r_1^{p_1} \ldots r_n^{p_n}.$$

The following Abelian theorem is valid:

THEOREM 4.1. *If $S(I)$ exists and*

$$|S_{q_1 \ldots q_n}| < C,$$

where C is an absolute constant, then $S(r)$ is uniformly convergent for $0 < r_j < 1$ and $\lim_{r\to I} S(r) = S(I)$. ($r \to I$ means $[r, 0] \to [I, 0]$.) See **(4)** for a proof.

Also a Tauberian theorem proved by Knopp **(7)** for double series may be extended to multiple series.

THEOREM 4.2 (Tauberian theorem). *Let the series $S(r)$ converge for each $[r, 0] \in D$, $r = [r_1, \ldots, r_n]$, and for these r let $|S(r)| < K$, where K is an absolute constant. If*

$$|c_{p_1\ldots p_n}|(p_1^2 + \ldots + p_n^2)^{\frac{1}{2}n} < M < \infty, \tag{2}$$

then $\lim_{r\to I} S(r) = S$ *implies* $S(I) = S$, *that is,*

$$\sum_{p_1=0}^{\infty} \cdots \sum_{p_n=0}^{\infty} c_{p_1\ldots p_n} = S. \tag{3}$$

In order to prove Theorem 4.2, Theorems 3 of § 3 and the proofs in § 4 of Knopp's paper must be proved for n-fold series ($n > 3$). Using condition (2) Theorem 3 has been extended to multiple series in **(18)**. Also by means of (2) the proofs in § 4 follow for multiple series. In addition see **(13)** where a similar condition is utilized for multiple series summed spherically.

On the other hand if we let $z \to [I, 0]$ along the set $[\rho I, 0]$, $0 < \rho < 1$, and sum series (3.14) by diagonals:

$$S_0(\rho) = \sum_{p=0}^{\infty} b_p \rho^p, \tag{4}$$

where

$$b_p = \sum_{p_1+\ldots+p_n=p} c_{p_1\ldots p_n}, \tag{4a}$$

then (4) is a simple series and the boundedness conditions on $S_{q_1\ldots q_n}$ and $S(r)$ can be omitted in Theorems 4.1 and 4.2 **(17)**. Abel's theorem reads *if $S_0(I)$ exists, then $S_0(\rho I)$ is uniformly convergent for $0 < \rho < 1$ and* $\lim_{\rho\to 1} S_0(\rho I) = S_0(I)$ and Theorem 4.2 becomes

Let the series $S_0(\rho I)$ converge for each $[\rho I, 0] \in D$, $(0 < \rho < 1)$. If $b_p = 0(1/p)$, then $\lim_{\rho\to 1} S_0(\rho I) = S_0$ *implies* $S_0(I) = S_0$.

2. *Cauchy's inequality and mean value theorem.* In the next paragraphs it is convenient to group the elements of the CONS $\{\phi_\nu\}$ of same degree, hence, let

$$\phi_1^{(p)}, \ldots, \phi_{M_p}^{(p)}$$

be the terms of degree p. Then for any $f \in L^2$ on D

$$f(z) = \sum_{p=0}^{\infty} \sum_{j=1}^{M_p} a_j^{(p)} \phi_j^{(p)}(z), \tag{5}$$

where the convergence is continuous and absolute for $z \in D$. Multiply (5) by $\bar{f}$ and integrate over the domain

$$D_\rho = [z | \rho^2 I - zz^* > 0, 0 < \rho < 1]. \tag{6}$$

Clearly $D_\rho \subsetneqq D$ for $0 < \rho < 1$. Since the convergence of (5) is uniform and absolute on $\bar{D}_\rho$

$$(7)\quad I(\rho) = V_\rho^{-1}\int_{D\rho}|f|^2 dW_z = V_\rho^{-1}\sum_{p=0}^{\infty}\sum_{j=1}^{M_p}\sum_{q=0}^{\infty}\sum_{k=1}^{M_q} a_j^{(p)}\bar{a}_k^{(q)}\int_{D\rho}\phi_j^{(p)}\bar{\phi}_k^{(q)}dW_z$$

(V_ρ being the volume of D). In the integral

$$I = \int_{D\rho}\phi_j^{(p)}\bar{\phi}_k^{(q)}dW_z$$

set $z = \rho w$. Then $D_\rho \to D$, $dW_z \to \rho^{2mn}dW_w$ (cf. § 3.1) and

$$\phi_j^{(p)}(z) = \sum_{\Sigma a_{jk}=p} \text{constant} \prod z_{jk}^{a_{jk}} \to \rho^p\phi_j^{(p)}(w).$$

Hence

$$I = \rho^{2mn+p+q}(\phi_j^{(p)}, \phi_k^{(q)}) = \rho^{2mn+p+q}\delta_{pq}\delta_{jk}.$$

Also $V_\rho = \rho^{2mn}V_0$ where V_0 is the volume of D. Thus (7) becomes

$$(8)\qquad I(\rho) = \frac{1}{V_\rho}\int_{D\rho}|f|^2 dW = \frac{1}{V_0}\sum_{p=0}^{\infty}\rho^{2p}\sum_{j=0}^{M_p}|a_j^{(p)}|^2.$$

From (8)

$$\frac{1}{V_0}\sum_{j=0}^{M_p}|a_j^{(p)}|^2 < (1/\rho^{2p})\max_{z\epsilon D\rho}|f(z)|^2.$$

Now according to a theorem proved by Hua **(6)** if f is analytic on the closed circular domain $\bar{D}_\rho$, then f attains its maximum modulus on the circular manifold

$$(9)\qquad B_\rho = [z|zz^{*\prime} = \rho^2 I].$$

Hence we get the Cauchy inequality:

$$(10)\qquad \frac{1}{V}\sum_{j=0}^{M_p}|a_j^{(p)}|^2 < (1/\rho^{2p})\max_{z\epsilon B\rho}|f(z)|^2.$$

THEOREM 4.3 (mean value theorem). *$I(\rho)$ defined by (7) is a monotone increasing function of $\rho (0 < \rho < 1)$ and* $\log I(\rho)$ *is a convex function of* $\log \rho$.

Proof. The monotonicity of $I(\rho)$ is obvious from (8). The proof of convexity is the same as the proof in **(17**, p. 174**)** for the one variable case.

3. *A mean value theorem over B_ρ.*

THEOREM 4.4. *In the case $n = m$ the integral*

$$(11)\qquad I_1(\rho) = \int_{B\rho}|f|^2 dV$$

is a monotone increasing function of ρ and $\log I_1(\rho)$ *is a convex function of* $\log \rho$.

Proof. Hua **(6)** has shown how to construct a set $\{\psi_\nu\}$ orthonormalized with respect to the inner product

$$(\psi_\nu, \psi_\mu)_B = \int_B \psi_\nu \bar{\psi}_\mu dV$$

from the CONS $\{\phi_\nu\}$ as follows. Since B is a circular space $(\phi_j^{(p)}, \phi_k^{(q)})_B = 0$ if $p \neq q$. Define the vector

$$z_p = (\phi_1^{(p)}, \ldots, \phi_{M_p}^{(p)}).$$

Then

$$(z_p', z_p)_B = ((\phi_j^{(p)}, \phi_k^{(p)})_B) = K_p$$

is a matrix of constants. Since $z_p'\bar{z}_p > 0$ if

$$z_p z_p^* = |\phi_1^{(p)}|^2 + \ldots + |\phi_{M_p}^{(p)}|^2$$

is positive, $K_p > 0$ and there exists a unitary matrix U such that $U^* K_p U = \Lambda$, where Λ is a diagonal matrix with positive elements on the diagonal. Now $\{y_p\}$, defined by

$$y_p = z_p \bar{U} = (\theta_1^{(p)}, \ldots, \theta_{M_p}^{(p)}),$$

is a CONS on D if $\{z_p\}$ is, since

$$((z_p \bar{U})', z_p U) = U^*(z_p', z_p)U = U^*((\phi_j^{(p)}, \phi_k^{(p)}))U = U^* I U = I.$$

Let

$$\psi_j^{(p)} = \theta_j^{(p)} / ||\theta_j^{(p)}||_B.$$

Then $\{\psi_j^{(p)}\}$ is an ONS with respect to integration over B and the orthogonal development (5) of $f \in L^2$ can be written as

$$f(z) = \sum_{p=0}^{\infty} \sum_{j=1}^{M_p} b_j^{(p)} \psi_j^{(p)}, \tag{12}$$

where

$$b_j^{(p)} = (f, \psi_j^{(p)}) / ||\psi_j^{(p)}||^2.$$

Multiply (12) by $\bar{f}$ and integrate over B_ρ. By a procedure similar to that in paragraph 3 we obtain the formula

$$I_1(\rho) = \int_{B_\rho} |f|^2 dV = \sum_{p=0}^{\infty} \rho^{2p} \sum_{j=1}^{M_p} |b_j^{(p)}|^2, \tag{13}$$

from which the conclusions of the theorem follow.

Note added in Proof. Recently it has come to my attention that Hua and Look **(21)** have proved that $F(z) \to f(u_o)$ as $z \to u_o$ in any manner. Further for continuous f on B, the solution F given by (3) is unique. They also consider Abel summability for continuous functions of the unitary group.

References

1. A. A. Albert, *Modern higher algebra* (Chicago, 1937).

2. S. Bergman, *The kernel function and conformal mapping*, Math. Surveys, No. V (1950).

3. S. Bochner, *Group invariance of Cauchy's formula in several variables*, Ann. Math., *45* (1944), 686–707.

4. T. J. I'A Bromwich and G. H. Hardy, *Some extensions to multiple series of Abel's theorem on continuity of power series*, Proc. London Math. Soc., *2* (1905), 161–189.
5. E. Cartan, *Sur les domains bornés homogènes de l'espace de n variables complexes*, Abh. Math. Sem. Univ. Hamburg, *11* (1936), 116–162.
6. L. K. Hua, *On the theory of functions of several complex variables*, I-III, Acta Math. Sinica, *2* (1953), 288–323; *5* (1955), 1–25, 205–242; (in Chinese), M.R., *17* (1956), p. 191. Also *Harmonic analysis of the classical domain in the study of analytic functions of several complex variables*, mimeographed notes (about 1956).
7. K. Knopp, *Limitierungs-Umkehrsätze für Doppelfolgen*, Math. Z., *45* (1939), 573–589.
8. D. B. Lowdenslager, *Potential theory and a generalized Jensen-Nevalinna formula for functions of several complex variables*, Jn. Math. Mech., *7* (1958), 207–218.
9. ——— *Potential theory in bounded symmetric homogeneous complex domains*, Ann. Math., *67* (1958), 467–484.
10. C. C. Macduffee, *The theory of matrices* (Ergebnisse der Mat. und ihrer Grenzgebiete [Chelsea, 1946]).
11. J. Mitchell, *On double Sturm-Liouville series*, Amer. J. Math., *65* (1943), 616–636.
12. ——— *An example of a complete orthonormal system and the kernel function in the geometry of matrices*, Proceedings of the Second Canadian Mathematical Congress (Vancouver, 1949), 155–163.
13. ——— *On the spherical summability of multiple orthogonal series*, Trans. Amer. Math. Soc., *71* (1951), 136–151.
14. ——— *Potential theory in the geometry of matrices*, Trans. Amer. Math. Soc., *79* (1955), 401–422.
15. ——— *Orthogonal systems on matric spaces*, West. Research Lab. Pub., Scientific Paper 60-94801-1-P2 (1956).
16. K. Morita, *On the kernel functions of symmetric domains*, Sci. Reports of Tokyo Kyoiku Daigaku Sec. A, *5* (1956), 190–212.
17. E. C. Titchmarsh, *The theory of functions* (2nd ed., Oxford University Press, 1939).
18. S. H. Tung, *Tauberian theorems for multiple series*, unpublished Master's thesis (The Pennsylvania State University).
19. H. Weyl, *The classical groups* (Princeton University Press, 1946).
20. A. Zygmund, *Trigonometrical series*, Monog. Mat., Tom V (Warsaw, 1935).
21. L. K. Hua and K. H. Look, *Theory of harmonic functions in classical domains*, Sci. Sinicd, 8, No. 10 (1959), 1032-1094.

The Pennsylvania State University

A THEOREM ON PARTIALLY ORDERED SETS, WITH APPLICATIONS TO FIXED POINT THEOREMS

SMBAT ABIAN AND ARTHUR B. BROWN

In this paper the authors prove Theorem 1 on maps of partially ordered sets into themselves, and derive some fixed point theorems as corollaries.

Here, for any partially ordered set P, and any mapping $f: P \rightarrow P$ and any point $a \in P$, a well ordered subset $W(a) \subset P$ is constructed. Except when $W(a)$ has a last element ξ greater than or not comparable to $f(\xi)$, $W(a)$, although constructed differently, is identical with the set A of Bourbaki **(3)** determined by a, f, and P_1: $\{x | x \in P, x \leqslant f(x)\}$.

Theorem 1 and the fixed point Theorems 2 and 4, as well as Corollaries 2 and 4, are believed to be new.

Corollaries 1 and 3 are respectively the well-known theorems given in **(1**, p. 54, Theorem 8, and Example 4**)**.

The fixed point Theorem 3 is that of **(1**, p. 44, Example 4**)**; and has as a corollary the theorem given in **(2)** and **(3)**.

The proofs are based entirely on the definitions of partially and well ordered sets and, except in the cases of Theorem 4 and Corollary 4, make no use of any form of the axiom of choice.

In what follows, "$a < b$" implies that a and b are distinct. Furthermore, we shall always deal with elements and subsets of a given partially ordered set P, and "lub T" will denote exclusively "the least upper bound of T in P"; that is, an upper bound z of T such that if s is any other upper bound of T, then $z < s$. The symbol "$\subset$" shall mean "is a subset (not necessarily proper) of."

Definition. Let P be a partially ordered set and f a mapping of P into P. For any $a \in P$, an ***a-chain*** C_r is a subset of P satisfying the following conditions:

(1) C_r is well ordered, with a as its first element and r as its last element;

(2) If $z \in C_r$ and $z \neq r$, then $f(z) \in C_r$, $z < f(z)$, and there exists no $y \in C_r$ for which $z < y < f(z)$;

(3) If T is a non-empty subset of C_r, then the least upper bound (in P) of T exists and is in C_r.

It will follow from Lemma 4 below that, for given P, f, and a, C_r is uniquely determined by r.

We designate by $W(a)$ the set of all $r \in P$ for which there exists an a-chain C_r having r as its last element. We note that (2) implies that $W(a) = \{a\}$ except when $a < f(a)$.

Received September 1, 1959.

Under the hypotheses of the Definition, we shall first prove the following lemmas.

LEMMA 1. *If $r \in W(a)$ and C_r is an a-chain with last element r, then $C_r \subset W(a)$.*

Proof. If $t \in C_r$, the set of all elements of C_r which are $\leqslant t$ is easily seen to be an a-chain, and hence $t \in W(a)$. Therefore the lemma is true.

LEMMA 2. *If $r \in W(a)$ and $r < f(r)$, then $f(r) \in W(a)$.*

Proof. The set $C_r \cup \{f(r)\}$ is obviously an a-chain, and hence $f(r) \in W(a)$.

LEMMA 3. *If $r, s \in W(a)$ and C_r is an a-chain with last element r, then either $s \in C_r$ or $r < s$.*

Proof. Let $T = C_r \cap C_s$. By (1), $T \neq \emptyset$ and hence, by (3), $z = \text{lub } T$ exists and $z \in T$. If $s \notin C_r$ then $z \neq s$. If also $z \neq r$ then, by (2), $z < f(z) \in T$, contrary to the fact that $z = \text{lub } T$. Hence $s \notin C_r$ implies that $z = r$, so that $r \in C_s$; and since $r \neq s$, we see by (1) that $r < s$. Since, by (1), $s \in C_r$ and $r < s$ cannot both hold, we infer the truth of Lemma 3.

LEMMA 4. *If $r \in W(a)$, there is just one C_r with last element r, namely the set of all elements of $W(a)$ which are $\leqslant r$.*

Proof. This follows from Lemmas 1 and 3.

THEOREM 1. *Let P be a non-empty partially ordered set, f a map of P into P, and a an arbitrary element of P. Then*

(4) *$W(a)$ is well ordered with a its first element.*

Moreover, if $\xi = \text{lub } W(a)$ exists, then

(5) *$W(a)$ is an a-chain with ξ its last element,*

and

$$\xi \nless f(\xi). \tag{6}$$

Proof. Let H be any non-empty subset of $W(a)$, and $r \in H$. Since $r \in H \cap C_r$, we see by (1) that $H \cap C_r$ has a first element, which, in view of Lemma 4, is the first element of H. Hence $W(a)$ is well ordered. By Lemma 1, $a \in W(a)$, and by (1), if $r \in W(a)$ then $a \leqslant r$. Thus we conclude that a is the first element of $W(a)$. Hence (4) is valid.

Next, assume $\xi = \text{lub } W(a)$ exists and let $W^* = W(a) \cup \{\xi\}$. We shall show that W^* is an a-chain. Since $W(a)$ is well ordered, W^* is well ordered too and thus (1) is satisfied for W^*, with a its first and ξ its last element. Now, let $z \in W^*$ and $z \neq \xi$. Then $z \in W(a)$ and $\{x|x \in W(a), z < x\} \neq \emptyset$. Sinec $W(a)$ is well ordered, z has an immediate successor r in $W(a)$, hence in W^*. By Lemma 4, z and r are the last two elements of C_r. Hence, by (2) applied to z as an element of C_r, we see that $f(z) = r$, so that (2) is satisfied for W^*. To prove (3) for W^*, let T be any non-empty subset of W^*. Obviously ξ is an

upper bound of T. If there is no element of $W(a)$ which exceeds every element of T then, in view of the well orderedness of $W(a)$, any upper bound of T is also an upper bound of $W(a)$ and hence is $\geqslant \xi$, which implies that $\xi = \text{lub } T$, and thus lub $T \in W^*$. If there is an element $r \in W(a)$ which exceeds every element of T, then $T \subset W(a)$ and, by Lemma 4, $T \subset C_r$. Hence, by (3), lub T exists and is in C_r, and therefore, by Lemma 1, lub $T \in W(a)$, so that again lub $T \in W^*$. Consequently (3) is satisfied for W^*. Therefore W^* is an a-chain with ξ its last element, which implies that $\xi \in W(a)$ and $W(a) = W^*$. Thus (5) is valid.

Now, suppose $\xi < f(\xi)$. By Lemma 2, $f(\xi) \in W(a)$, so that (5) is contradicted. Therefore $\xi \nless f(\xi)$. Thus (6) is valid, and Theorem 1 is proved.

THEOREM 2. *Let P be a partially ordered set in which*

(7) lub *of every non-empty well ordered subset $W \subset P$ exists.*

Let f be a map of P into P such that f is isotone, that is,

(8) *for every two elements $x, y \in P$ with $x \leqslant y$, we have $f(x) \leqslant f(y)$;*

and

(9) *there exists an element $a \in P$ with $a \leqslant f(a)$.*

Then there exists at least one $\xi \in P$ such that $\xi = f(\xi)$. In fact, $\xi =$ lub *$W(a)$ is such an element.*

Proof. If $a = f(a)$, the conclusion is obvious. Now suppose $a < f(a)$.

Consider the set $W(a)$, where a is the element referred to in (9). By (4) and (7), $\xi = \text{lub } W(a)$ exists, and hence by (5), $W(a) = C_\xi$. By (9) and Lemma 2 we see that $f(a) \in W(a)$, and therefore $a < \xi$. Since $W(a)$ is an a-chain and $W(a) - \{\xi\}$ is non-empty, we infer from (3) that $\theta = \text{lub}[W(a) - \{\xi\}]$ is in $W(a) = C_\xi$. According as $\theta = \xi$ or $\theta < \xi$, we have

$$\xi = \text{lub } [W(a) - \{\xi\}] \tag{10}$$

or

(11) ξ is the immediate successor of θ in $W(a)$.

If (10) holds, take any element $z \in [W(a) - \{\xi\}]$. Then $z < \xi$, and by (8), $f(z) \leqslant f(\xi)$. By (2), $z < f(z)$. Consequently $z < f(\xi)$ and therefore $f(\xi)$ is an upper bound for $[W(a) - \{\xi\}]$, and thus, by (10), $\xi \leqslant f(\xi)$.

If (11) holds, by (2), $f(\theta) = \xi$. Also, since $\theta < \xi$, by (8), $f(\theta) \leqslant f(\xi)$, so that again $\xi \leqslant f(\xi)$.

Since $\xi \leqslant f(\xi)$, we see from (6) that $\xi = f(\xi)$. Thus Theorem 2 is proved.

Remark. An alternative proof of Theorem 2 can be given by considering the set $\{x | x \in P, x \leqslant f(x)\}$ and using Theorem 1.

COROLLARY 1. *Let f be any isotope map of a non-empty complete lattice L into itself. Then $\xi = f(\xi)$ for some $\xi \in L$.*

Proof. In view of Theorem 2, we need only verify (9). Choose $a =$ the greatest

lower bound of L. Then clearly (9) is valid, and Corollary 1 follows from Theorem 2.

COROLLARY 2. *Let P be a partially ordered set in which*

(12) *every non-empty well ordered subset $W \subset P$ which is bounded above has a* lub.

Let f be an isotone map of P into P and let there exist two elements $a, b \in P$ such that

(13) $$a \leqslant f(a) \leqslant f(b) \leqslant b.$$

Then there exists $\xi \in P$ such that $\xi = f(\xi)$ and $a \leqslant \xi \leqslant b$. In fact, $\xi = \operatorname{lub} W(a)$ is such an element.

Proof. Let $Q = \{x | x \in P, a \leqslant x \leqslant b\}$. Since f is isotone, we see by (13) that if $x \in Q$, then $a \leqslant f(a) \leqslant f(x) \leqslant f(b) \leqslant b$. Hence f maps Q into Q. Moreover, since Q is bounded above by b, we see from (12) that (7) is valid for Q. Therefore the hypotheses of Theorem 2 are satisfied by Q, f, and a. Thus from Theorem 2 we infer the validity of Corollary 2.

COROLLARY 3. *If f is an isotone map of a conditionally complete lattice into itself and if $a \leqslant f(a) \leqslant f(b) \leqslant b$, then $\xi = f(\xi)$ for some ξ with $a \leqslant \xi \leqslant b$.*

THEOREM 3. *Let P be a non-empty partially ordered set in which*

(14) lub *of every non-empty well ordered subset $W \subset P$ exists.*

Let f be a map of P into P such that

(15) $$\text{for every } x \in P, \qquad x \leqslant f(x).$$

Then there exists at least one $\xi \in P$ such that $\xi = f(\xi)$. In fact, for every $a \in P$, $\xi = \operatorname{lub} W(a)$ is such an element.

Proof. Consider an a-chain $W(a) \subset P$. By (4) and (14), $\xi = \operatorname{lub} W(a)$ exists. By (15) and (6), $\xi = f(\xi)$. Thus Theorem 3 is proved.

In the following a generalization of Corollary 2 is proved with the help of the axiom of choice.

THEOREM 4. *Let P be a partially ordered set in which*

(16) lub *of every non-empty well ordered subset which is bounded above exists.*

Let g be a map of P into P such that, for every two elements $x, y \in P$,

(17) $$\text{if } g(x) < g(y), \text{ then } x < y;$$

and, for $x, y, s \in P$,

(18) $$\text{if } g(x) \leqslant s \leqslant g(y), \text{ then } g^{-1}(s) \neq \emptyset.$$

Furthermore, let f be an isotone map of P into P, and let there exist $a, b \in P$, with $a < b$, satisfying

$$g(a) \leqslant f(a) \qquad \text{and} \qquad f(b) \leqslant g(b).$$

Then the reexists at least one $\xi \in P$ such that $a \leqslant \xi \leqslant b$ and $f(\xi) = g(\xi)$.

Proof. If $f(a) = g(a)$ or $f(b) = g(b)$, the conclusion is obvious. Hence we may assume that

(19) $$g(a) < f(a) \quad \text{and} \quad f(b) < g(b).$$

Consider the set $\{S_i\}$ of all non-empty subsets $S_i \subset P$ such that there exists $s_i \in P$ with $g^{-1}(s_i) = S_i$. Clearly, $\{S_i\} \neq \emptyset$. By the axiom of choice, there exists a function φ mapping $\{S_i\}$ into P, such that $\varphi(S_i) \in S_i$. Hence

(20) $$g\varphi g^{-1}(s_i) = s_i.$$

We observe also that, in view of (17),

(21) if $s_i < s_j$, then every element of $g^{-1}(s_i)$ < every element of $g^{-1}(s_j)$.

We shall show now that the function

(22) $$h = \varphi g^{-1} f$$

maps the set $Q = \{x | x \in P,\ a \leqslant x \leqslant b\}$ into itself. If $x \in Q$, then, since f is isotone, by (19) we have

(23) $$g(a) < f(a) \leqslant f(x) \leqslant f(b) < g(b),$$

and hence by (18) we see that $g^{-1}[f(x)] \neq \emptyset$. By (21) and (23) we find, $a < \varphi g^{-1}[f(x)] < b$. Hence, by (22), $h(x) \in Q$. Taking $x = a$, we infer also that

(24) $$a < h(a).$$

Furthermore, since f is isotone, if $x \leqslant y$ then $f(x) \leqslant f(y)$, and from (21) we infer that $\varphi g^{-1}[f(x)] \leqslant \varphi g^{-1}[f(y)]$, so that by (22) h is isotone on Q.

From (24) we see that a and h satisfy (9) on Q. Also, since Q is bounded above by b, we see from (16) that Q satisfies (7).

Hence Q and h satisfy the hypotheses of Theorem 2, and consequently there exists $\xi \in Q$ such that $h(\xi) = \xi$. Applying g to each side we have, by (22)

$$g\varphi g^{-1}[f(\xi)] = g(\xi),$$

and thus, by (20),

$$f(\xi) = g(\xi).$$

This completes the proof.

Corollary 4. *If in Theorem* 4 *instead of condition* (17) *we assume that* g *is isotone, then the conclusion of Theorem* 4 *remains valid provided* P *is a simply ordered set.*

References

1. G. Birkhoff, *Lattice theory* (A.M.S. Coll. Publ., *25*, rev. ed.; New York, 1948).
2. N. Bourbaki, *Théorie des ensembles* (Paris, 1956), chapter III, p. 48, Example 6b.
3. ——— *Sur le Théorème de Zorn*, Archiv der Mathematik, *2* (1949–50), 434–437.

University of Pennsylvania
and
Queens College

ARITHMETIC LINEAR TRANSFORMATIONS AND ABSTRACT PRIME NUMBER THEOREMS

S. A. AMITSUR

1. Introduction. Shapiro and Forman have presented in **(4)** an abstract formulation of prime number theorems which includes the various prime number theorems; for primes in arithmetic progressions, for prime ideals in ideal classes etc. The methods of proofs are "elementary" and follow closely Shapiro's proof for the primes in arithmetic progression (for reference see bibliography in **(4)**).

The author has followed in **(1)** some ideas of Yamamoto **(5)** on arithmetic linear transformations to introduce a symbolic calculus in dealing with arithmetic functions. This calculus proved to be very useful in unifying many of the "elementary" proofs in the behaviour of arithmetic functions. In **(6)** Yamamoto has extended his theory to ideals in algebraic number fields, and with this extension the symbolic calculus of **(1)** can be extended to cover the abstract case of prime number theorem in countable free abelian groups as discussed in **(4)**. Furthermore, a more careful study of the behaviour of certain "remainders" yields a more general result in the direction given by Beurling **(3)**.

Shapiro and Forman have considered the following situation. Let G be a free abelian group on a countable number of generators p_i $(i = 1, 2, \ldots,)$. $N{:}G \to \mathfrak{N}$ be a homomorphism of G into the multiplicative group of all integers $\mathfrak{N}$, with the kernel G' such that G/G' is finite. If H is a generic class of G/G', and w is an integral word in G, then the main result of **(4)** is deriving from the condition

$$\sum_{\substack{Nw\leqslant x\\ w\epsilon H}} 1 = c_H x + R_H(x);\, c_H \geqslant 0,\ \sum c_H > 0 \tag{1.1}$$

a "prime number theorem" for the class H:

$$\pi_H(x) = \sum_{\substack{Np\leqslant x\\ p\epsilon H}} 1 = d_H \frac{x}{\log x} + o\left(\frac{x}{\log x}\right). \tag{1.2}$$

A complete analysis of the coefficient of d_H was given in **(4)** for the case $R_H(x) = O(x^\theta)$ with $1>\theta \geqslant 0$. The methods developed in the present paper will show that the same results are valid even if $R_H(x) = O(x/\log^\gamma x)$ with $\gamma > 2$. A result of a similar nature, though in a completely different situation, was given by Beurling **(3)** for $\gamma > \frac{3}{2}$.

It is quite surprising that for $\gamma > 3$ (and in certain cases for $\gamma > 4$) the methods and the results of **(1)** can be carried over to the abstract case almost

Received April 27, 1959.

without any change whereas $3 > \gamma > 2$ involves many refinements of the methods and of the main "elementary proof" of **(1)**. In fact, some of the equivalent forms of the prime number theorem cannot be proved by our methods for $4 > \gamma > 2$; though others can be shown for these values of γ relatively easily, their classical proofs of implying the prime number theorem breaks down if $\gamma < 3$.

We take this opportunity to present also the symbolic calculus of **(1)** in a more general and in what we hope is a simplified form. An application is also given to show (by elementary methods) that $\zeta(1 + it) \neq 0$ for $t \neq 0$ and that $\sum p^{-1+it}$ converges.

2. The semi-group W and its characters. In the present context we prefer to consider the semi-group W of all integral words in G, and similarly $W' = W \cap G'$. In this way the group K of all characters of G/G' **(4)** is replaced by a finite group of characters of W. To be more precise, we assume the following:

Let W be a free abelian multiplicative semi-group generated by a countable number of generators p_i. Let N be a homomorphism of W into the multiplicative semi-group $\mathfrak{N}$ of all positive integers, that is, Nw is an integer and $N(w_1w_2) = Nw_1 . Nw_2$.

Let K be a finite group of characters of W. By a character $\chi \in K$, we mean a homomorphism of W into the complex numbers. The unit $\chi_0 \in K$ is defined as $\chi_0(w) = 1$ for all $w \in W$. Multiplication in K is given by:

$$(\chi\eta)(w) = \chi(w)\eta(w). \tag{2.1}$$

Let K be a finite group of order h, then it follows readily by (2.1) that $\chi(w)$ is an hth root of unity. Furthermore, each $w \in W$ determines a character of K by setting $\bar{w}(\chi) = \chi(w)$. Thus the mapping $w \to \bar{w}$ is a homomorphism of W into the group $\hat{K}$ of all characters of K. Let W' be the kernel of this map, that is,

$$W' = \{w; w \in W, \chi(w) = 1 \text{ for all } \chi \in K\}.$$

This readily implies that W/W' is a finite group of order $\leqslant$ order of K = order of $\hat{K} = h$. Now the classes H of W/W' are determined by the group of characters K; that is, u, v belong to the same class H if and only if $\chi(u) = \chi(v)$ for all $\chi \in K$, or in other words if and only if $\bar{u} = \bar{v}$. On the other hand, K is readily seen to induce a group of characters on the finite group W/W', and from the definition of the classes of the latter it follows that different characters of K induce different characters of W/W'. Consequently h = order of $K \leqslant$ order of W/W'. Combining this with the previous result, we obtain:

PROPOSITION 1. *W/W' is a finite group of order h, and K can be considered as the group of all characters of W/W'.*

In many cases the converse situation is preferred. Namely, given $W' \subseteq W$ such that W/W' is finite, we define K to be the group of all characters of W/W',

and then $\chi(w)$ is defined to be $\chi(H)$ where $w \in H$ the class in W/W'. In any case, we shall always use the notation $\chi(H)$ and $\chi(w)$ for the same character χ.

Now the standard relation between characters yields:

$$\sum_{H} \chi(H)\eta(H) = \begin{cases} 0 & \text{if} \quad \chi \neq \bar{\eta} \\ h & \text{if} \quad \chi = \eta \end{cases} \tag{2.2}$$

$$\sum_{\chi} \chi(u)\chi(v) = \begin{cases} 0 & \text{if } u, v \text{ belong to different classes of } W/W' \\ h & \text{if } u, v \text{ belong to the same class.} \end{cases} \tag{2.3}$$

Next we assume that for any class

$$B_H(x) = \sum_{\substack{Nw<x \\ w\in H}} 1 = C_H x + O(x/\log^\gamma x);\, C_H > 0,\ \sum C_H > 0. \tag{2.4}$$

We define

$$\psi_H(x) = \sum_{\substack{Np^t<x \\ p^t\in H}} \log Np;\ \pi_H(x) = \sum_{\substack{Np<x \\ p\in H}} 1. \tag{2.5}$$

Analogous to the results of Shapiro and Forman **(4)**, we shall show that the character can be distributed into three classes Γ_1, Γ_2, Γ_3. Γ_1 will contain all character for which $A_\chi = \sum_H \chi(H)\, C_H \neq 0$, Γ_2 and Γ_3 will be defined later in §8. Our first result is:

Theorem A. *If $\gamma > 2$, then*

$$\sum_{Np^t<x} \chi(p^t) \log Np = \begin{cases} x + o(x) & \text{if} \quad \chi \in \Gamma_1 \\ o(x) & \text{if} \quad \chi \in \Gamma_2 \\ -x + o(x) & \text{if} \quad \chi \in \Gamma_3 \text{ and } \gamma > 3. \end{cases}$$

Let $U = \{w; \chi(w) = 1 \text{ for all } \chi \in \Gamma_1\}$, and $U^* = \{w; w \in U, \chi(w) = 1 \text{ for all } x \in \Gamma_3\}$. Then $W' \subseteq U^* \subseteq U$ and as in **(4**, Theorem 3.1**)**:

Theorem B. *If* (2.4) *holds then:*

$$\pi_H(x) = d_H \frac{x}{\log x} + o\left(\frac{x}{\log x}\right),$$

if (a) $\Gamma_3 = \emptyset$, $\gamma > 2$, *where:*

$$d_H = \begin{cases} 0 & \text{for } H \notin U/W' \\ h^{-1} \text{ order } \Gamma_1 & \text{for } H \in U/W' \end{cases}$$

or (b) $\Gamma_3 \neq \emptyset$, $\gamma > 3$ *where:*

$$d_H = \begin{cases} 0 \text{ for } H \notin U/W' \text{ or } H \in U^*/W' \\ 2h^{-1} \text{ order } \Gamma_1 \quad \text{for } H \in U/W' \text{ or } H \notin U^*/W'. \end{cases}$$

3. The ring C(W). Let $C(W)$ be the set of all complex valued functions of W. As in **(5**, p. 42**)**, $C(W)$ is a ring with respect to the addition

$$(f + g)(w) = f(w) + g(w);\ w \in W, \tag{3.1}$$

and the convolution:

$$(3.2)\qquad (f*g)(w) = \sum_{uv=w} f(u)g(v).$$

We shall also use the ordinary multiplication:

$$(3.3)\qquad (fg)\,(w) = f(w)g(w).$$

The ring $C(W)$ is in fact a commutative ring with the unit ϵ defined.

$$(3.4)\qquad \epsilon(1) = 1,\ \epsilon(w) = 0 \text{ for } w = 1 \text{ (the unit of } W).$$

As in the classical case $W = \mathfrak{N}$ the integers, **(1, 5)** it is easily shown that the invertible function $f \in C(W)$ are those for which $f(1) \neq 0$, and in this case $f^{-1}(w)$ is defined by induction on the length of the words.

$$(3.5)\quad f^{-1}(1) = 1/f(1);\, f^{-1}(w) = -\left[\sum_{u|w} f^{-1}(u)f(wu^{-1})\right]\Big/ f(1),\ u \neq w.$$

Let E be the "one" function defined $E(w) = 1$ for all $w \in W$, then its inverse $E^{-1} = \mu_w = \mu$ is the Mobius function for W:

$$(3.6)\qquad \begin{aligned} &\mu(w) = (-1)^r \text{ if } w \text{ is the product of } r \text{ distinct generators} \\ &\mu(p) = 1, \text{ and zero otherwise.} \end{aligned}$$

A function f is said to be *multiplicative* if:

$$(3.7)\qquad f(uv) = f(u)f(v) \text{ for } (u, v) = 1$$

where $(u, v) = 1$ means that u, v have no common divisor $\neq 1$ in W. If (3.7) holds for all u, v without any restriction, then we say that f is *factorable* or f is a *character*. Another type of functions which we meet are the *additive* functions which satisfy:

$$(3.8)\qquad f(u\,v) = f(u) + f(v).$$

In the general case of arbitrary semi-group W as in the case of the integer **(5)** we have:

PROPOSITION 2. *If f is a character, then the mapping: $g \to gf$ is an isomorphism of $C(W)$ into itself. In particular: $(g*h)f = (gf)*(hf)$.*

If f is an additive function, then the mapping: $g \to gf$ is a derivation of $C(W)$. In particular: $(g*h)f = (gf)*h + g*(hf)$.

Let N be the homomorphism of W into the semi-group of all integers $\mathfrak{N}$. We shall refer to Nw as the *norm* of w. Since N is a homomorphism, N is a character, and consequently the log-function L, defined thus:

$$(3.9)\qquad L(w) = \log Nw$$

is an additive function. Thus, it follows from Proposition 2 that

$$(3.10)\qquad (f*g)\,L = fL*g + f*gL \text{ for all } f, g \in C(W).$$

We shall use the notation L^m to mean $L^m(w) = \log^m (Nw)$. With the aid of $\mu = \mu_w$ we define as in **(6**, p. 44**)** the Mangolt-function $\Lambda = \Lambda_w = \mu * L$ the Selberg-function $\Lambda_2 = \mu * L^2$ and higher types $\Lambda_m = \mu * L^m$. We recall that

(3.11) $\Lambda(p^e) = \log Np$ and $\Lambda(w) = 0$ if $w \neq p^e$ for a generator $p \in W$.

(3.12) $\Lambda_2(p^e) = (2e - 1) \log Np$; $\Lambda_2(p^e q^f) = 2 \log Np \log Nq$;
$\Lambda_2(w) = 0$ for $w \neq p^e q^f$.

To every $f \in C(W)$ we define an arithmetic function $Nf \in C(\mathfrak{N})$ by setting

(3.13) $$(Nf)(n) = \sum_{Nw=n} f(w) \qquad \text{for every integer } n > 0,$$

and if there are no $w \in W$ satisfying, $Nw = n$ then we set $(Nf)(n) = 0$.

Thus $(NE)(n)$ is the number of elements of W whose norm is n. It is not difficult to show

THEOREM 1. *The mapping $f \to Nf$ is a homomorphism of $C(W)$ into the ring of all arithmetic function $C(\mathfrak{N})$.*

4. The ring of arithmetic linear transformations. Let F be the linear space of all complex valued functions $\Phi(x)$ defined for all real $x \geqslant 1$. To each $f \in C(W)$, we make correspond (as in **(1, 5)**) a linear transformation S_f of F, defined by

(4.1) $$(S_f\Phi)(x) = \sum f(w)\Phi(x/Nw); \; \Phi \in F \text{ and all } x \geqslant 1.$$

The following is then easily verified.

PROPOSITION 3.

$$S_{f+g} = S_f + S_g; \; cS_f = S_{cf}; \; S_{f*g} = S_f S_g.$$

That means that the correspondence: $f \to S_f$ is a homomorphism of $C(W)$ into the ring of all linear transformations of F.

Definition (4.1) is valid for all semi-groups, in particular for $W = \mathfrak{N}$ (the integers) where in the semi-group of integer the norm is to be the identity map. Then clearly we have, by (3.13),

PROPOSITION 4.

$$S_f \Phi = S_{Nf}\Phi.$$

For practical purposes we prefer to substitute for S_f a different operator I_f defined by

(4.2) $$(I_f\Phi)(x) = \sum_{Nw \leqslant x} \frac{f(w)}{Nw} \Phi\left(\frac{x}{Nw}\right) = (S_{fN^{-1}}\Phi)(x)$$

where

$$(fN^{-1})(w) = f(w)/Nw.$$

As we remarked above, N and therefore also N^{-1} are characters of W, hence it follows readily by Proposition 2 that Propositions 3 and 4 will hold also for I_f. For further references we formulate this result in the following proposition which includes also an additional simple fact.

PROPOSITION 5.
$$I_f + I_g = I_{f+g}; \quad cI_f = I_{cf}; \quad I_f I_g = I_{f*g}; \quad I_f\Phi = I_{Nf}\Phi,$$
and
$$I_f(\Phi \log x) = \log x \,.\, I_f\Phi - I_{fL}\Phi.$$

5. The space $\mathfrak{L}$. In the present section we extend the formalism introduced in **(1)** to cover the general case dealt with in the present paper.

Let $\mathfrak{L}$ be the space of all polynomials $\phi(\log x) = \sum a_\nu \log^\nu x$ in the function $\log x$. We introduce the formal derivation $D = d/d \log x$ with all its positive and negative powers by writing

$$D^m \log^n x = (n)_m \log^{n-m} x \text{ for all } n \geqslant m,\ n \geqslant 0, \tag{5.1}$$
$$= 0 \qquad\qquad \text{if } m > n,$$

where $(n)_m = n!/(n-m)!$ if $n \geqslant m$ and $n \geqslant 0$; m can be positive or negative. Thus, D^0 is the identity. For completeness we set $(n)_m = 0$ if $m > n$. Now D^m acts on $\Phi(\log x)$ by setting: $D^m(\sum a_\nu \log^\nu x) = \sum (\nu)_m a_\nu \log^{\nu-m} x$.

Let $\alpha_{-p}, \alpha_{-p+1}, \ldots, \alpha_0, \ldots,$ be a sequence of complex numbers, then the symbol

$$F(D) = \sum_{\nu=-p}^{\infty} \alpha_\nu D^\nu$$

will be considered as a linear operator on $\mathfrak{L}$, by putting

$$F(D) \log^n x = \sum_{\nu=-p}^{\infty} \alpha_\nu D^\nu \log^n x = \sum_{\nu=-p}^{n} (n)_\nu \alpha_\nu \log^{n-\nu} x. \tag{5.2}$$

Let $f \in C(W)$, $F(D)$ be as above. Then we denote by $R_n(x; f, F)$ the remainder element defined by the relation

$$I_f \log^n x = F(D) \log^n x + R_n(x; f, F). \tag{5.3}$$

That is, in view of (5.2)

$$R_n(x; f, F) = \sum_{Nw<x} \frac{f(w)}{Nw} \log^n \frac{x}{Nw} - \sum_{\nu=-p}^{n} (n)_\nu \alpha_\nu \log^{n-\nu} x. \tag{5.4}$$

As in **(1)** we shall write

$$I_f = F(D) + O(\varphi_n) \tag{5.5}$$

to mean

$$R_n(x; f, F) = O(\varphi_n(x)), \text{ for all } n \geqslant 0.$$

The notations $R_n(x)$, $R_n(x; f)$ and $R_n(f)$ will replace $R_n(x; f, F)$ when no confusion will be involved.

For further references we fix .

$$F(D) = \sum_{\nu=-p}^{\infty} \alpha_\nu D^\nu, G(D) = \sum_{\mu=-q}^{\infty} \beta_\mu D^\mu.$$

The following is easily verified.

THEOREM 2. (1) $\alpha R_n(x;f, F) + \beta R_n(x;g,G) = R_n(x;\alpha f + \beta g, \alpha F + \beta G)$
(2) $R_n(x;fL, -F') = \log x \cdot R_n(x;f, F) - R_{n+1}(x;f, F)$

where $F' = \sum \nu\alpha_\nu D^{\nu-1}$ *is the formal derivative with respect to* D.

The proof of (2) follows as in the proof of (4) of (**1**, Theorem 4.1).

Another simple result which is of great importance in the present paper is

THEOREM 3.

$$R_n(x;g*f, GF) = I_g R_n(x;f, F) + \sum_{j=0}^{n+p} (n!/j!)\alpha_{n-j} R_j(x;g, G)$$
$$- \sum_{t=0}^{q-1} \sum_{s=0}^{t+1} \alpha_{n+t+s}\beta_{-s}(n!/t!) \log^t x.$$

This will be used mainly in the following form, (noting that $\alpha_{-p} \neq 0$)

$$R_{n+p}(x;g,G) = cI_g R_n(x;f, F) + \sum_{j=0}^{n+p-1} c_j R_j(x;g,G) + \sum_{t=0}^{q-1} c_{n,t} \log^t x + dR_n(x;g*f, GF). \tag{5.6}$$

for some constants c, c_j, $c_{n,t}$, d. In both formulas if $n + p < 0$, the term containing $R_j(x;g,G)$ does not appear, and if $q - 1 < 0$ the last term is not to be considered.

We note also that $G(D)F(D)$ is the formal product of the two power series in D and not the product of the operator G and F; the two products are not always equal as can be seen by: $1 = (D^{-1}D)1 \neq D^{-1}(D1) = 0$.

Proof.

$$I_{g*f} \log^n x = I_g(I_f \log^n x) = I_g\left[\sum_{\nu=-p}^{n} (n)_\nu \alpha_\nu \log^{n-\nu} x + R_n(x;f, F)\right]$$
$$= I_g R_n(x;f, F) + \sum_{\nu=-p}^{n} (n)_\nu \alpha_\nu I_g \log^{n-\nu} x$$
$$= I_g R_n(x;f, F) + \sum_{\nu=-p}^{n} (n)_\nu \alpha_\nu R_{n-\nu}(x;g, G)$$
$$+ \sum_{\nu=-p}^{n} \sum_{\mu=-q}^{n-\nu} (n)_\nu \alpha_\nu (n-\nu)_\mu \beta_\mu \log^{n-\nu-\mu} x$$
$$= A + B + C = (GF) \log^n x + R_n(x;g*f, GF).$$

The terms A, B appear in the statement of Theorem 3 (by setting $j = n - \nu$)

and if $n + p < 0$, we do not get B. To complete the proof of Theorem 3 we have to compare

$$[G(D)F(D)] \log^n x = \sum_{k=-(p+q)}^{n} \left(\sum_{\nu+\mu=k} \alpha_\nu \beta_\mu \right) (n)_k \log^{n-k} x$$

with

$$C = \sum_{\nu=-p}^{n} \sum_{\mu=-q}^{n-\nu} (n)_\nu (n-\nu)_\mu \alpha_\nu \beta_\mu \log^{n-\nu-\mu} x = \sum_{k=-(p+q)}^{n} \left(\sum_{\nu+\mu=k}' \alpha_\nu \beta_\mu \right) (n)_k \log^{n-k} x,$$

which is obtained by setting $\nu + \mu = k$. The difference between the two is that in $(GF) \log^n x$, the sum ranges over all $\nu \geqslant -p, \mu \geqslant -q$, whereas in the second sum it ranges only over: $n \geqslant \nu \geqslant -p$, $n - \nu \geqslant \mu \geqslant -q$. Comparing the two we observe that they have common range as long as min $(n, k+q) \geqslant \nu \geqslant -p$ with $k = \nu + \mu \leqslant n$. Thus the terms for which $k + q \geqslant \nu > n$ show that:

$$C - (GF)\log^n x = - \sum_{k=n-q+1}^{n} \left(\sum_{\nu+\mu=k}'' \alpha_\nu \beta_\mu \right) (n)_k \log^{n-k} x$$

$$= \sum_{t=0}^{q-1} \sum_{s=1}^{t+1} (n!/t!) \alpha_{n+t+s} \beta_{-s} \log^t x,$$

since in $\sum''$, $\nu > n$; hence, the last form is obtained by setting $t = n - k$ and $s = -\mu$, as then $\nu = k - \mu = n + t + s$. (If $q = 0$, this term does not appear, since $k + q \leqslant n$.)

The relation (5.6) is very useful in computing $R_n(x; f^{-1}, F^{-1})$ by induction, since it provides us with a recursive formula for $R_n(x; f^{-1}, F^{-1})$ as will be used later.

Another formula for $R_n(g*f)$ has been obtained in **(2)** following the Dirichlet hyperbola method for summation. This result has been proved only for the integers (Theorem 1 of **(2)**), and we formulate it here for the semi-group W, but the two results are equivalent as is readily seen by the equality $I_f = I_{Nf}$, which leads to the relation: $R_n(x; f, F) = R_n(x; Nf, F)$ for all $f \in C(W)$. We quote that result in the following theorem.

THEOREM 4. *Let* $yz = x$; $1 \leqslant y \leqslant x$ *then*

$$R_n(x; g*f, GF) = \sum_{Nw \leqslant y} \frac{g(w)}{Nw} R_n\left(\frac{x}{Nw}; f, F\right) + \sum_{Nw \leqslant z} \frac{f(w)}{Nw} R_n\left(\frac{x}{Nw}; g, G\right)$$

$$- \sum_{j=0}^{n} \binom{n}{j} R_{n-j}(y; g, G) R_j(z; f, F)$$

$$+ \sum_{i=1}^{p} \sum_{j=1}^{i} \frac{n!}{(i-j)!(n+j)!} \alpha_{-i} R_{n+j}(y; gG) \log^{i-j} z +$$

$$\sum_{i=1}^{q} \sum_{j=1}^{i} \frac{n!}{(i-j)!(n+j)!} \beta_{-i} R_{n+j}(z; f, F) \log^{i-j} y,$$

and the respective terms do not appear if the power series of $F(D)$, $G(D)$ *do not have negative powers of* D(that is, $p < 0$ or $q < 0$).

The following lemma will be used extensively in §6.

LEMMA 1. *Let*

$$\Phi(x) = \sum_{Nw<x} f(w),$$

let $\varphi(x)$ *be a differentiable function. Then*

$$\sum_{Nw<x} f(w)\varphi(Nw) = \Phi(x)\varphi(x) - \int_1^x \Phi(t)d\varphi(t),$$

or more generally

$$\sum_{y<Nw<x} f(w)\varphi(Nw) = \Phi(x)\varphi(x) - \Phi(y)\varphi(y) - \int_y^x \Phi(t)d\varphi(t).$$

This lemma follows immediately from (7, Theorem 421, p. 346) noting that

$$\sum_{Nw<x} f(w) = \sum_{n<x} (Nf)(n).$$

6. Approximating I_f. In the following two sections we consider functions $f \in C(W)$ with properties

$$S_f 1 = \sum_{Nw<x} f(w) = \alpha x + O(x/\log^\gamma x), \qquad \gamma = 1 + \delta > 0, \tag{6.1}$$

$$S_{|f|} 1 = \sum_{Nw<x} |f(w)| = Ax + O(x/\log^\gamma x), \tag{6.2}$$

or the weaker condition:

$$\sum_{Nw<x} |f(w)| = O(x). \tag{6.2*}$$

For later applications we shall introduce the assumption

(6.3) f^{-1} exists in $C(W)$ and $|f^{-1}(w)| < K|f(w)|$ for some $K > 0$, and all $w \in W$

These function will satisfy

PROPOSITION 6.

$$I_f 1 = \sum_{Nw<x} (Nw)^{-1} f(w) = \alpha \log x + \alpha_0 + \rho(x), \qquad \rho(x) = O(\log^{-\delta} x) \tag{6.4}$$

$$S_{fL} 1 = \sum_{Nw<x} f(w) \log Nw = \alpha x \log x - \alpha x + \alpha + O(x \log^{-\delta} x) \tag{6.5a}$$

$$S_{fL^2} 1 = \sum_{Nw<x} f(w) \log^2 Nw = \alpha x \log^2 x - 2\alpha x \log x + 2\alpha x - 2\alpha + O(x \log^{1-\delta} x). \tag{6.5b}$$

The proof follows immediately from (6.1) by applying Lemma 1. We observe also that, since $\delta = 1 - \gamma > 0$.

$$\alpha_0 = \alpha + \int_1^\infty O(t^{-1}\log^{-\gamma} t)dt;\ \rho(t) = O(\log^{-\gamma} x) + \int_x^\infty O(t^{-1}\log^{-\gamma} t)dt = O(\log^{-\delta} x).$$

In what follows we determine an approximation of I_f assuming only the validity of (6.4), and to simplify results, we assume henceforth that $\delta = \gamma - 1$ is not an integer.

From Lemma 1 we obtain, for $n > 0$,

$$\begin{aligned}
I_f \log^n x &= \sum_{Nw<x} (Nw)^{-1} f(w) \log^n (x/Nw) = -\int_1^x (\alpha \log t + \alpha_0 + \rho(t))\, d\log^n (x/t) \\
&= (n+1)^{-1}\alpha \log^{n+1} x + \alpha_0 \log^n x - \sum_{\nu=1}^{n} \binom{n}{\nu}(-1)^\nu \log^{n-\nu} x \int_1^x \rho(t) d\log^\nu t \\
&= (n+1)^{-1}\alpha \log^{n+1} x + \alpha_0 \log^n x + \sum_{1\leq\nu<\delta} \frac{(-1)^{\nu-1}}{(\nu-1)!}\int_1^\infty \rho(t) t^{-1}\log^{\nu-1} t\, dt \cdot \frac{n!}{(n-\nu)!}\log^{n-\nu} x \\
&\quad + \sum_{\nu<\delta} (-1)^\nu \binom{n}{\nu} \log^{n-\nu} x \cdot \int_x^\infty \rho(t) d\log^\nu t - \sum_{\nu>\delta} (-1)^\nu \binom{n}{\nu} \log^{n-\nu} x \int_1^x \rho(t) d\log^\nu t.
\end{aligned}$$

This is true since, for $\nu < \delta$,

$$\int_1^\infty \rho(t) d\log^\nu t = \nu \int_1^\infty t^{-1}\rho(t)\log^{\nu-1} t\, dt < \infty$$

as $\rho(t) = O(\log^{-\delta} x)$. If $n < \delta$, we disregard the last term. Put

$$(6.6)\qquad \begin{cases} F(D) = \sum_{\nu=-1}^{\infty} \alpha_\nu D^\nu \text{ with } \alpha_{-1} = \alpha,\ \alpha_0 \text{ as given in (6.4)} \\ \alpha_\nu = O \text{ for } \nu > \delta, \\ \alpha_\nu = \dfrac{(-1)^{\nu-1}}{(\nu-1)!}\displaystyle\int_1^\infty \frac{\rho(t)\log^{\nu-1} t}{t} dt \text{ for } 1 \leq \nu < \delta. \end{cases}$$

Thus we have obtained that $I_f \log^n x = F(D)\log^n x + R_n(x; f, F)$ where

$$(6.7)\quad R_n(x;f,F) = \sum_{\nu<\delta} (-1)^\nu \binom{n}{\nu} \log^{n-\nu} x \int_x^\infty \rho(t) d\log^\nu t - \sum_{\nu>\delta} (-1)^\nu \binom{n}{\nu} \log^{n-\nu} x \int_1^x \rho(t) d\log^\nu t$$

and for the case $n = 0$, we have clearly by (6.4)

$$(6.7^*)\qquad R_0(x;f,F) = \rho(x).$$

If $n < \delta$ we can obtain a better form for R_n, namely

$$R_n(x; f, F) = \sum_{\nu=1}^{n} (-1)^\nu \binom{n}{\nu} \log^{n-\nu} x \cdot \int_x^\infty \rho(t) d \log^\nu t \tag{6.8}$$
$$= \int_x^\infty \rho(t) d \log^n (xt^{-1}) = (-1)^n \int_0^\infty \rho(xe^u)\, du^n,$$

where the latter is obtained by setting $u = -\log (xt^{-1})$.

If (6.3) holds for all $\delta > 0$ (that is, (6.1) is valid for all $\gamma > 0$) then we define α_ν by the integral of (6.6) for all $\nu > 1$. Furthermore, it follows readily from (6.8) that, for all $n < \delta$,

$$R_n(x; f, F) = \pm \int_0^\infty \rho(xe^u) du^n = O(\log^{n-\delta} x). \tag{6.9}$$

Thus we have

THEOREM 5. *If*

$$\sum_{Nw<x} (Nw)^{-1} f(w) = \alpha_{-1} \log x + \alpha_0 + O(\log^{-\delta} x)$$

for all $\delta > 0$ *then* $I_f = F(D) + O(\log^{-\delta} x)$ *for all* $\delta > 0$, *and* $F(D)$ *is as given in* (6.6).

In many cases we can obtain a better bound for $R_n(x; f, F)$.

If $\rho(x) = O(x^{-\vartheta})$, $\vartheta > 0$, then one readily obtains from (6.8) that

$$R_n(x; f, F) = O\left(\int_0^\infty x^{-\vartheta} e^{-\vartheta u} du^n\right) = O(x^{-\vartheta}).$$

COROLLARY. *If*

$$\sum_{Nw<x} f(w) = \alpha x + O(x^{1-\vartheta})$$

then $I_f = F(D) + O(x^{-\vartheta})$.

Now, if (6.9) is valid only for a bounded δ, then we can only show

THEOREM 6. $R_n(x; f, F) = O(\log^{n-\delta} x)$ *with* $F(D)$ *as given in* (6.6).

Indeed, for $\nu < \delta$,

$$\int_x^\infty \rho(t) d \log^\nu t = O(\log^{\delta-1} x)$$

and for $\nu > \delta$

$$\int_1^x \rho(t) d \log^\nu t = O(\log^{\delta-1} x).$$

Thus our theorem follows immediately from (6.7).

Applying Theorem 2 to this approximation of I_f yields

THEOREM 7.

$$I_{fL} = -\sum_{\nu=-1}^{\infty} \nu \alpha_\nu D^{\nu-1} + O(\log^{n+1-\delta} x),$$

and generally

$$I_{fL^m} = (-1)^m F^{(m)}(D) + O(\log^{n+m-\delta} x),$$

where

$$F(D) = \sum_{\nu=-1}^{\infty} \alpha_\nu D^\nu$$

is given in (6.6) *and* $F^{(m)}(D)$ *denotes the* mth *formal derivative of* $F(D)$ *with respect to* D.

The coefficients $\alpha_{-1} = \alpha$, α_0, α_1, A are defined in (6.2) and (6.6). We have to deal separately with the following cases.

Case I $\alpha_{-1} \neq 0$ (this implies that $A \neq 0$).

Case II $\alpha_{-1} = 0$, $\alpha_0 \neq 0$, and $A \neq 0$.

Case III $\alpha_{-1} = \alpha_0 = 0$ (this will imply that $A \neq 0$).

Case IV $\alpha_{-1} = 0$, $\alpha_0 \neq 0$ and $A = 0$

Our first purpose is to show that

THEOREM 8.

$$I_{f^{-1}} = F^{-1}(D) + O(1) + \begin{cases} O(\log^{n-\delta} x) \text{ in Cases I, IV} \\ O(\log^{n+1-\delta} x) \text{ in Case II} \\ O(\log^{n+2-\delta} x) \text{ in Case III.} \end{cases}$$

and the four cases contain all possible conditions on the coefficients.

We shall need the following lemma.

LEMMA 2. *Let* $h(w) > 0$ *satisfying*

$$S_h 1 = \sum_{Nw<x} h(w) = O(x).$$

Let $g(x) = O(\log^r x)$ *be a non-negative bounded function in finite intervals, then* $I_h g = O(1) + O(\log^{r+1} x)$. *If* $S_h 1 = O(x/\log^r x)$, $r > 1$ *then*: $I_h g = O(1) + O(\log^r x)$.

Indeed, let $|g(x)| < K \log^r x$ for $x > a$, then

$$|I_h g| < \sum_{xa^{-1}<Nw<x} (Nw)^{-1} h(x)|g(x/Nw)| + K \sum_{Nw<xa^{-1}} (Nw)^{-1} h(w) \log^r (x/Nw)$$

$$< \sup_{1<t<a} |g(t)|(xa^{-1})^{-1} \cdot O(x) + K(xa^{-1})^{-1} \log^r a \cdot O(x)$$

$$- K \int_1^{xa^{-1}} O(t) d[t^{-1} \log^r xt^{-1}] = O(1) + O(\log^{r+1} x)$$

as can readily be obtained by substituting $u = x/t$ in the integral.

The second result follows similarly if we observe that

$$\int_1^{xa^{-1}} O(t \log^{-r} t) d[t^{-1} \log^r xt^{-1}] < K \int_1^s d[t^{-1} \log^r xt^{-1}] +$$

$$+ L \int_s^{xa^{-1}} t \log^{-r} t d[t^{-1} \log^r xt^{-1}]$$

for some $1 < s < a^{-1}$, and some constants $K, L > 0$. Clearly, the first integral is $O(\log^r x)$ and the second is:

$$= O\left(\int_s^{xa^{-1}} t^{-1} \log^{-r} t \log^r (xt^{-1}) dt\right) = O(\log^r x).$$

As a special case, if $f(w)$ satisfies (6.2) and (6.3) and $g(x)$ is as above, then we have

COROLLARY 2.

$$I_{f^{-1}} g = O(I_{|f|} |g|) = O(1) + O(\log^{r+1} x) \qquad \text{if } A \neq 0,$$

$$I_{f^{-1}} g = O(I_{|f|} |g|) = O(1) + O(\log^r x) \qquad \text{if } A = 0.$$

Indeed, by (6.3) it follows that $|f^{-1}(w)| < k|f(w)|$ for some $K > 0$. Thus, $I_{f^{-1}} g = O(I_{|f|}|g|)$ and the rest follows by the preceding lemma.

Next we prove

PROPOSITION 7. *If f satisfies (6.1)–(6.3) and $\delta > 1$ (that is, $\gamma > 2$), then one of the coefficients $\alpha_{-1}, \alpha_0, \alpha_1$ of (6.6) is not zero. Furthermore, if $\alpha_{-1} = \alpha_0 = 0$ then $A \neq 0$.*

For, let $\alpha_{-1} = \alpha_0 = 0$, then from Theorem 6 we deduce that $I_f \log x = \alpha_1 + O(\log^{1-\delta} x)$. Applying $I_{f^{-1}}$ on both sides and using the preceding corollary, we obtain

$$\log x = I_{f^{-1}} I_f \log x = \alpha, I_{f^{-1}} 1 + O(\log^{2-\delta} x) + O(1) = \alpha_1 I_{f^{-1}} 1 + o(\log x)$$

since $2 - \delta < 1$, and this can only be true if $\alpha_1 \neq 0$. Moreover, in this case, it follows in view of (6.3) that:

$$|\alpha_1^{-1} \log x + o(\log x)| < |I_{f^{-1}} 1| < K I_{|f|} 1 = AK \log x + O(1)$$

hence $A \neq 0$. This proves also that the four cases described in Theorem 8 cover all possible cases. (Note, that at this point only in Case IV we assumed $\gamma > 2$.)

Remark. If in (6.2), we assume that $A = 0$, then clearly $\alpha_{-1} = 0$, since $|I_f 1| < I_{|f|} 1$, and in this case it follows that $\alpha_0 \neq 0$. The latter is then true even for $\delta > 0$, since we can use the better bound given in Corollary 2 for

$$I_{f^{-1}} O(\log^{1-\delta} x).$$

So that if $\alpha_0 = 0$ we would have

$$\log x = I_{f^{-1}} I_f \log x = \alpha_1 I_{f^{-1}} 1 + I_{f^{-1}} O(\log^{1-\delta} x) = \alpha_1 I_{f^{-1}} 1 + O(\log^{1-\delta}),$$

which implies that $\alpha_1 \neq 0$ even for $\delta > 0$. But then

$$|\alpha_1^{-1} \log x + o(\log x)| = |I_{f^{-1}}1| < I_{|f|}1 = O|\log^{-\delta}x)$$

which is a contradiction. Hence, $\alpha_0 \neq 0$.

We are now in position to prove Theorem 8.

Case I. Since $\alpha = \alpha_{-1} \neq 0$, $F^{-1}(D) = \alpha^{-1}D + \ldots$ and therefore

$$I_{f^{-1}}1 = F^{-1}(D)1 + R_0(x; f^{-1}, F^{-1}).$$

That is,

$$R_0(x; f^{-1}, F^{-1}) = I_{f^{-1}}1.$$

To evaluate this element, consider the following.

$$\begin{aligned} 1 = S_{f^{-1}}S_f1 &= S_{f^{-1}}[\alpha x + O(x \log^{-\gamma}x)] = \alpha x I_{f^{-1}}1 + xI_{f^{-1}}O(\log^{-\gamma}x) \\ &= \alpha x R_0(x; f^{-1}, F^{-1}) + xO(\log^{-\gamma+1}x) + xO(1). \end{aligned}$$

Since $S_f x = xI_f1$, now since $-\delta = 1 - \gamma$, we have shown that $R_0(x; f^{-1}, F^{-1}) = O(1) + O(\log^{-\delta}x)$. We complete the proof of this case by induction on n. Observing that:

$$O = R_n(x; \epsilon, 1) = R_n(x; f^{-1}*f, F^{-1}F)$$

we obtain by (5.6), (where $p = 1$) in view of Theorem 6 and Corollary 2,

$$\begin{aligned} R_{n+1}(x; f^{-1}, F^{-1}) &= O[I_{f^{-1}}R_n(x; f, F)] + O\left(\sum_{j=0}^{n} |R_j(x; f^{-1}, F^{-1})|\right) \\ &= \theta[I_{|f|}O(\log^{n-\delta}x)] + \sum_{j=0}^{n} O(\log^{j-\delta}x) + O(1) = O(\log^{n+1-\delta}x) + O(1), \end{aligned}$$

which completes the proof of this case.

Case II. The proof follows by a similar application of (5.6). In this case, $p = 0$ and we need no special method for computing. As we have by (5.6)

$$R_n(x; f^{-1}, F^{-1}) = O[I_{f^{-1}}R_n(x; f, F)] + O\left(\sum_{j=0}^{n-1} |R_j(x; f^{-1}, F^{-1})|\right)$$

where for $n = 0$, the sum does not appear. Thus using again an induction, together with Corollary 2 and Theorem 6, we obtain that

$$R_n(x; f^{-1}, F^{-1}) = O(\log^{n+1-\delta}x) + O(1).$$

Incidentally, this provides the proof for Case IV also, since there $A = 0$ and we can use the better approximation

$$I_{f^{-1}}R_n(x; f, F) = O(\log^{n-\delta}x) + O(1)$$

which will yield in Case IV

$$R_n(x; f^{-1}, F^{-1}) = O(\log^{n-\delta}x).$$

Case III. Again we use the same procedure, but here $p = -1$. So that

$$R_{n-1}(x;f^{-1}, F^{-1}) = O[I_{f^{-1}}R_n(x;f, F)] + O\left(\sum_{j=0}^{n-2} |R_j (x;f^{-1}, F^{-1})|\right)$$

and we thus obtain $R_{n-1}(x;f^{-1}, F^{-1}) = O(\log^{n+1-\delta}x)$, which completes the proof of Theorem 8.

It follows now readily from Theorem 3 that

THEOREM 9. For $m \geqslant 1$:

$$I_{f^{-1}*fL^m} = (-1)^m F^{-1}(D)F^{(m)}(D) + O(1) + \begin{cases} O \log^{n+m+1-\delta}x) \text{ in Cases I–III} \\ O(\log^{n+m-\delta}x) \text{ in Case IV.} \end{cases}$$

Indeed, Theorem 3 implies

$$R_n(x;f^{-1}*fL^m) = I_f R_n(fL^m) + O\left(\sum_{j=0}^{n+p} |R_j (f^{-1})|\right) + O(1)$$

where D^{-p} is the first power of D appearing in $F^{(m)}(D)$, and $O(1)$ has to be added only if $\alpha_{-1} = \alpha_0 = 0$, since then $F^{-1}(D) = \alpha_{-1}{}^{-1}D^{-1} + \dots$ (that is, $q = 1$).

Since $F(D)$ has at most one negative power of D, that is, D^{-1}, the mth derivative may have the lowest power $D^{-(m+1)}$, thus $p \leqslant m + 1$. Furthermore, in view of Corollary 2 and Theorem 7,

$$I_{f^{-1}}R_n(fL^m) = O(\log^{n+m+1-\delta}x) + O(1).$$

The other terms can get at most to this power, by Theorem 8, which proves Cases I–III. In Case IV $p = 0$ and we can apply the better bound of Corollary 2 to yield the required result.

In particular this leads to

COROLLARY 3. If

$$\sum_{Nw<x} f(w) = \alpha x + O(x \log^{-\delta}x)$$

for all $\delta > 0$ and then

$$I_{f^{-1}*fL^m} = (-1)^m F^{-1}(D)F^{(m)}(D) + O(1).$$

This includes the known results **(1)** about $I_\mu, I_\Lambda, I_{\Lambda_2}$ where μ, Λ, Λ_2 are the Mobius', Mangoll's, and Selberg's function for the integers, respectively. More applications will be given later.

7. Approximating characters. Let f be a character on W, then the preceding results can be further refined in the direction of the "elementary proofs" developed in **(1)**. This can be achieved relatively easily following the proofs of Theorem 9.1 and 9.2 of **(1)**—only if we assume that $\gamma > 3$ where γ is given in (6.1). We shall outline the proofs of this fact later.

In the present section we want to obtain results which will give us the proof of Theorem A and B even for $\gamma > 2$. We will be able to obtain the result that $\sum f(w)\Lambda(w) = x + o(x)$ if $\gamma > 2$ in most cases, whereas the relation $\sum f(w)Nw^{-1}\Lambda(w) = \log x + c + o(1)$ will be obtained only for $\gamma > 3$.

In the rest of this section we assume that

(A) f is a character satisfying (6.1), (6.2), and $A \neq 0$, $\gamma = 1 + \delta > 2$. Since f is a character, it follows by Proposition 2 that $f^{-1}(w) = f(w)\mu_W(w)$ which shows that f satisfies also (6.3). For these characters we show

PROPOSITION 8.

$$(7.1) \qquad S_{f\Lambda}1 = \sum_{Nw<x} f(w)\Lambda(w) = O(x)$$

$$(7.2) \qquad S_{|f|\Lambda_2}1 = \sum_{Nw<x} |f(w)|\Lambda_2(w) = 2x\log x + O(x) + O(x\log^{2-\delta}x),$$

(Selberg's formula)

$$(7.3) \qquad \left|\sum_{x<Nw<tx} f(w)\Lambda(w)\right| < 2(t-1)x + o(x) \text{ as } (t, x) \to (1, \infty).$$

Proof. It follows from Proposition 2, that since f is a character,

$$(7.4) \qquad f^{-1} = f\mu; f\Lambda = f(\mu * L) = f^{-1} * fL \text{ and } f\Lambda_2 = f(\mu * L^2) = f^{-1} * fL^2.$$

As the mapping $g \to gL$ is a derivation in $C(W)$, we have: $(\mu * L)L = \mu L * L + \mu * L^2 = -(\mu * \mu * L) * L + \mu * L^2 = -(\mu * L)^2 + (\mu * L^2)$. Hence:

$$(7.5) \qquad f\Lambda_2 = f\Lambda^2 + f\Lambda L.$$

Now $|f|$ is also a character, hence it follows by (6.5a) that:

$$\begin{aligned} x^{-1}\sum |f(w)|\Lambda(w) &= I_{|f|\Lambda}x^{-1} = I_{|f|^{-1}}I_{|f|L}x^{-1} = I_{|f|^{-1}}x^{-1}S_{|f|L}1 \\ &= I_{|f|^{-1}}x^{-1}[Ax\log x - Ax + A + O(x\log^{-\delta}x)] \\ &= AI_{|f|^{-1}}\log x - AI_{|f|^{-1}}1 + x^{-1}AS_{|f|^{-1}}1 + I_{|f|^{-1}}O(\log^{-\delta}x) \\ &= O(1) + O(\log^{1-\delta}x) = O(1) \end{aligned}$$

which follows immediately by Corollary 2 and Theorem 8. This gives the proof of (7.1). The proof of (7.2) follows similarly by use of (6.5b). Namely

$$\begin{aligned} x^{-1}\sum |f(w)|\Lambda_2(w) &= I_{|f|^{-1}}(I_{|f|L^2}x^{-1}) \\ &= I_{|f|^{-1}}[A\log^2 x - 2A\log x + 2A + O(\log^{1-\delta}x)] \\ &= 2\log x + O(1) + O(\log^{2-\delta}x), \end{aligned}$$

since by Theorem 2 $I_{|f|^{-1}}\log^2 x = (A^{-1}D + \ldots)\log^2 x + O(\log^{2-\delta}x)$.

The proof of (7.3) follows now by standard methods from (7.2) and (7.5). That is

$$0 < \sum_{x<Nw\le xt} |f(w)|\,\Lambda(w) < \sum_{x<Nw\le tx} |f(w)|\,\Lambda(w)\log Nw \log^{-1} x$$
$$< \log^{-1} x \sum_{x<Nw\le tx} \Lambda_2(w) = \log^{-1} x(2tx\log tx - 2x\log x) + O(x\log^{-1} x) +$$
$$O(x\log^{1-\delta} x) = (2t-1)x + o(x).$$

The "elementary proofs" lie in the following refinement of (**2**, Theorem 4).

THEOREM 10. *Let $g(w) \in C(W)$ be a non-negative function satisfying*

(g1) $$\sum_{Nw\le x} g(w) = Mx\log^n x + o(x\log^n x); \qquad M > 0, n \geqq 1.$$

Let $h(x)$ be a real a complex-valued function which satisfies

(h1) $$h(x) = O(1)$$

(h2) $$\sum_{\nu\le x} \nu^{-1} h(\nu) = O(1)$$

(h3) $$h(tx) - h(x) = o(1) \quad \cdot s \quad (t, x) \to (1, \infty).$$

Then the condition

(g2) $$|h(x)|\log^{n+1} x < \frac{n+1}{M} \sum_{Nw\le x} \frac{g(w)}{Nw} \left| h\left(\frac{x}{Nw}\right)\right| + o(\log^{n+1} x)$$

implies

$$h(x) = o(1).$$

This theorem has been given in (**2**, Theorem 4) with the condition

$$\sum (Nw)^{-1} g(w) = a\log^{n+1} x + b\log^n x + o(\log^n x)$$

which is stronger than (g1), since (g1) implies only that

(7.6) $$I_g 1 = \sum (Nw)^{-1} g(w) = [Mx\log^n x + o(x\log^n x)]x^{-1}$$
$$+ \int_1^x [Mt\log^n t + o(t\log^n t)]t^{-2}dt = (n+1)^{-1} M\log^{n+1} x + o(\log^{n+1} x).$$

To prove this theorem, we first observe that for given $t > 1$, we can find x_t such that, for x, y satisfying $xy^{-1} > x_t$, the following is valid:

(7.7) $$\sum_{y\le x/Nw\le yt} (Nw)^{-1} g(w) > C\log^n(xy^{-1}) \quad \text{for some } C > 0.$$

Indeed, choose δ (to be fixed later) then there is x_δ such that for $x > x_\delta$, the absolute value of the error term in (g1) is $< \delta\log^n x$. Then it follows by (g1) that

$$\sum_{y\le x/Nw\le yt} (Nw)^{-1} g(w) \geqq yx^{-1} \sum g(w)$$
$$\geqq yx^{-1}[Mxy^{-1}\log^n xy^{-1} - Mx(yt)^{-1}\log^h x(yt)^{-1} - 2\delta xy^{-1}\log^n xy^{-1}]$$
$$\geqq M(1-t^{-1})\log^n xy^{-1} + Mt^{-1}[\log^h xy^{-1} - \log^n x(yt)^{-1}] - 2\delta\log^n xy^{-1}$$
$$\geqq M(1 - t^{-1} - 2\delta)\log^n xy^{-1},$$

and (7.7) is true if we choose $C = M(1 - t^{-1} - 2\delta) > 0$, which can be fulfilled as $t > 1$.

By the standard method of Selberg's proof (**1**, Theorem 6.1 and **2**, Theorem 4) one can obtain the following.

(7.8) Given $\Delta > 0$, there exists x_Δ, $T > t > 1$ such that for $x > x_\Delta$, there is y, $x \leqslant y < yt \leqslant xT$ with the property that for all $y \leqslant z \leqslant yt$, $|h(z)| < \Delta$.

We turn now to the proof of Theorem 10, which contains only a more careful repetition of the proof of (**2**, Theorem 4).

Let $\limsup |h(x)| = A$. If $A > 0$, choose $\Delta = \frac{1}{2}A$ and fix x_Δ, $T > t > 1$ satisfying (7.8). For this given t we choose x_δ to satisfy (7.7). Now for given $\epsilon > 0$, let $|h(x)| < A + \epsilon$ for all $x > X_\epsilon$.

Denote by y_i the element y given in (7.8) for $x = T^i > x_\Delta$, that is, $T^i < y_i < y_i . t < T^{i+1}$ and put $\xi = \log x_0 / \log T$ where $x_0 = \max(x_\epsilon, x_\Delta)$, and $\eta = \log(x x_\delta^{-1})/\log T$. Thus for each $\xi < i < \eta$, $T^i > x_0$ and $xT^{-i} > x_\delta$.

It follows now by (9.2) in view of (7.7), (7.8), and (7.6) that

$$\begin{aligned} |h(x)| \log^{n+1} x &\leqslant (n+1)M^{-1} \sum_{x(Nw)^{-1} \leqslant x_0} (Nw)^{-1} g(w) |h(x/Nw)| \\ &\quad + (n+1)M^{-1}(A+\epsilon) \sum_{xx_0^{-1} \leqslant x(Nw)^{-1} \leqslant x} (Nw)^{-1} g(w) \\ &\quad + \sum_{\xi < i < \eta} \sum_{y_i < (Nw)^{-1} x < y_i t} [\Delta - (A+\epsilon)](Nw)^{-1} g(w) \\ &\leqslant K \sum_{xx_0^{-1} \leqslant Nw \leqslant x} (Nw)^{-1} g(w) + (n+1)M^{-1}(A+\epsilon)[M(n+1)^{-1} \log^{n+1} x \\ &\quad + o(\log^{n+1} x)] + [\Delta - (A+\epsilon)]C \sum_{\xi < i < \eta} \log^n (xT^{-i-1}), \end{aligned}$$

since $\Delta - (A + \epsilon) < 0$ and $xy_i^{-1} > xT^{-i-1}$. Now, the first term is $\leqslant Kx_0 x^{-1}(Mx \log^n x + o(x \log^n x)) = o(\log^{n+1} x)$. For the third term we have by Lemma 1,

$$\begin{aligned} \sum_{\xi < i < \eta} \log^n (xT^{-1} \cdot T^{-i}) &= -[\xi] \log^n xT^{-\xi-1} + [\eta] \log^n xT^{-\eta-1} \\ &\quad - \int_\xi^\eta [u] d \log^n xT^{-u-1} = (n+1)^{-1} \log^{-1} T \log^{n+1} x + o(\log^{n+1} x) \end{aligned}$$

as follows immediately by standard method of replacing $[u]$ by u and noting that $T^\xi = x_0$, $T^\eta = xx_\delta^{-1}$.

Thus

$$|h(x)| \leqslant A + \epsilon + (\Delta - A - \epsilon)C/M \log T + o(1).$$

As $x \to \infty$ with $|h(x)| \to A$ we get $A \leqslant A + \epsilon + (\Delta - A - \epsilon)C'$, $C' > 0$. But this cannot be true for all $\epsilon > 0$ since $(\Delta - A)C' < 0$. This contradiction leads to the conclusion that $A = 0$.

We apply now Theorem 10 to the following function: $g(w) = |f(w)|\Lambda_2(w)$ and

$$h(x) = x^{-1} \sum_{Nw<x} f(w)\,\Lambda(w) + \begin{cases} -1 & \text{in Case I} \\ 0 & \text{in Case II} \\ +1 & \text{in Case III} \end{cases} \tag{7.9}$$

since $2 - \delta < 1$, it follows by (7.2) that $|f(w)|\Lambda_2(w)$ satisfies (g1). Clearly, (7.1) means that $h(x)$ given in (7.9) satisfies (h1). Condition (h3) follows by (7.1) and (7.3),

$$|h(tx) - h(x)| < \left| \sum_{Nw<x} f(w)\,\Lambda(w)(t^{-1} - 1)x^{-1} \right| + \left| \sum_{x<Nw<tx} f(w)t^{-1}x^{-1}\Lambda(w) \right|$$

$$< K(t^{-1} - 1) + 2(t - 1)t^{-1} + o(1) = o(1) \quad \text{as} \quad (t, x) \to (1, \infty).$$

To obtain (h2) we put $\sigma = 1$ in case I, $\sigma = 0$ for case II, and $\sigma = -1$ in Case III:

$$\left| \sum_{\nu<x} \nu^{-1}h(\nu) \right| = \left| \sum_{\nu<x} \nu^{-2} \sum_{Nw<\nu} f(w)\,\Lambda(w) - \sigma \sum_{\nu<x} \nu^{-1} \right|$$

$$= \left| \sum_{Nw<x} f(w)\,\Lambda(w) \cdot \sum_{Nw<\nu<x} \nu^{-2} - \sigma \log x + O(1) \right|$$

$$= \left| \sum f(w)\,\Lambda(w[(Nw)^{-1} - x^{-1} + O(Nw^{-2})] - \sigma \log x + O(1) \right.$$

$$< \textstyle\sum_1 + \sum_2 + \sum_3 + O(1),$$

where

$$\textstyle\sum_2 = \left| x^{-1} \displaystyle\sum_{Nw<x} f(w)\,\Lambda(w) \right| = O(1) \text{ by (7.1), and}$$

$$\textstyle\sum_3 = \left| \displaystyle\sum_{Nw<x} (Nw)^{-2}|f(w)|\,\Lambda(w) \right| < \displaystyle\sum_{Nw<x} |f(w)| \log Nw(Nw)^{-2} = O(1),$$

as follows immediately by (6.5b).

We can conclude from Theorem 9 that $\sum_1 = |I_{f\Lambda} - \sigma \log x| = O(1) + O(\log^{2-\delta} x)$ which is $O(1)$ if $2 - \delta < 0$, that is, $\delta > 2$ or $\gamma > 3$. From this we can conclude Theorem A for $\gamma > 3$. To obtain our result for $\gamma > 2$ we need a refinement of Theorem 9, which we can carry out only in the following form.

THEOREM 11. *If f is a character satisfying* (6.1) *and* (6.3) *with $A \neq 0$, then*

$$I_{f\Lambda} = -F^{-1}(D)F'(D) + O(1) + \begin{cases} O(\log^{n-\delta} x) & \text{in Case I} \\ O(\log^{n+1-\delta} x) & \text{in Case II} \\ O(\log^{n+2-\delta} x) & \text{in Case III.} \end{cases}$$

Before proceeding with the proof of this theorem, we observe that with these results it follows now that $\sum_1 = O(1)$ if $\delta > 1$ in Cases I and II, and only in Case III we have to assume that $\delta > 2$. To complete the proof of our first main theorem, we establish

THEOREM 12. *If f is a character satisfying* (6.1) *and* (6.2) *and* $A \neq 0$, *then*

$$\sum_{Nw<x} f(w)\,\Lambda(w) = \begin{cases} x + o(w) & \text{in Case I and } \gamma > 2, \\ o(x) & \text{in Case II and } \gamma > 2, \\ -x + o(x) & \text{in Case III and } \gamma > 3. \end{cases}$$

We still have to prove the validity of (g2). Indeed by (6.5a)

$$I_f h(x) = I_f(x^{-1}S_{f\Lambda}1) = x^{-1}S_f S_{f\Lambda}1 - \sigma I_f 1$$
$$= x^{-1}S_{fL}1 - \sigma I_f 1 = \begin{cases} O(\log^{-\delta} x) - \alpha_0 & \text{in Cases II and III} \\ O(\log^{-\delta} x) - \alpha_1 - \alpha_0 & \text{in Case I} \end{cases}$$

since $I_f x^{-1} = x^{-1}S_f$ (as operators) and $S_{f\Lambda} = S_{f^{-1}}S_{fL}$. Hence if $a = -\alpha_0$ or $a = \alpha_{-1} - \alpha_0$,

$$I_{f\mu} \log x\, I_f h(x) = aI_{f\mu} \log x + I_{f\mu}O(\log^{1-\delta} x)$$
$$= O(1) + O(\log^{2-\delta} x) = o(\log x),$$

by Theorem 8 and Corollary 2. We now proceed similarly to (**1**, Lemma 6.1)

$$I_{f\mu} \log x\, I_f = I_{f\mu}(I_f \log x + I_{fL}) = \log x + I_{f\Lambda}.$$

It follows therefore that

$$|h(x)|\log x - I_{|f|\Lambda}|h(x)| < |(\log x + I_{f\Lambda})h(x)| = |I_{f\mu}\log x\, I_f h(x)| = o(\log x).$$

As in (**1**, Lemmas 6.3 and 6.5) we obtain

$$(\log^2 x - I_{|f|\Lambda_2})|h(x)| = (\log x + I_{|f|}\ )(\log x - I_{|f|}\ )|h(x)|$$
$$< (\log x + I_{|f|}\ )o(\log x) = o(\log^2 x).$$

That is,

$$|h(x)|\log^2 x < I_{|f|\Lambda_2}|h(x)| + o(\log^2 x)$$

which proves (g2), after verifying easily as in (**1**, Lemma 6.3) with the aid of Theorem 11 that $I_{|f|\Lambda}o(\log x) = o(\log^2 x)$.

We return now to the proof of Theorem 11. From Lemma 2 and (7.1) it follows that

$$I_{|f|\Lambda}O(\log^r x) = O(1) + O(\log^{r+1} x). \tag{7.9}$$

The proof is similar to the proof of Theorem 8. It follows by (6.5a) that

$$\alpha x \log x - \alpha x + \alpha + O(x \log^{-\delta} x) = S_{fL}1 = S_{f\Lambda}S_f 1$$
$$= S_{f\Lambda}[\alpha x + o(x/\log^{1+\delta} x)] = \alpha x I_{f\Lambda}1 + xI_{f\Lambda}O(\log^{-1-\delta} x)$$
$$= \alpha x I_{f\Lambda}1 + xO(\log^{-\delta} x) + O(x).$$

Thus, if $\alpha \neq 0$, we have: $I_{f\Lambda}1 = \log x + O(1) + O(\log^{-\delta} x)$ which yield $R_0(x; f\Lambda, -F'F^{-1}) = O(1) + O(\log^{-\delta} x)$.

It follows now from the relation $fL = f{*}f\Lambda$ and by (5.6) and Theorem 7 using induction that

$$\begin{aligned} R_{n+1}(x; f\Lambda, -F^{-1}F') &= cI_{f\Lambda}R_n(x; f, F) + O\Big(\sum_{j=0}^{n} |R_j\,(x; f\Lambda, -F'F^{-1})|\Big) \\ &\qquad + O(1) + R_n(x; f\Lambda{*}f, -F^{-1}F'\cdot F) \\ &= I_{f\Lambda}O(\log^{n-\delta}x) + O(1) + R_n(x; fL, -F') + O(\log^{n-\delta}x) \\ &= O(\log^{n+1-\delta}x) + O(1), \end{aligned}$$

which prove the first case of Theorem 11.

The other cases follow as in Theorem 8:

$$\begin{aligned} R_{n+p}(x; f\Lambda, -F^{-1}F') &= cI_{f\Lambda}R_n(x; f, F) + O\Big(\sum_{j=0}^{n+p-1} |R;\,(x; f\Lambda, -F^{-1}F')|\Big) \\ &\qquad + O(1) + R_n(x; fL, -F') = O(\log^{n+1-\delta}x) + O(1). \end{aligned}$$

In Case II, $p = 0$ and in Case III, $p = -1$, which readily imply by induction the other two cases of Theorem 11.

We conclude this section with the last case $\alpha = A = 0$ (which implies $\alpha_0 \neq 0$). Here we do not have to use Theorem 10. The proof of (7.1) which leads to (7.9) holds in this case, and consequently, Theorem 11 (Case II) is also valid. Writing

$$h(x) = x^{-1}\sum_{Nw<x} \Lambda(w) = x^{-1}S_{f\Lambda}1.$$

As in the first part of the proof of Theorem 12, we obtain

$$\begin{aligned} h(x)\log x + I_{f\Lambda}h(x) &= I_{\mu f}\log x\, I_f(x^{-1}S_{f\Lambda}1) = I_{\mu f}x^{-1}\log x S_{fL}1 \\ &= I_{\mu f}O(\log^{1-\delta}x) = O(1) + O(\log^{2-\delta}x). \end{aligned}$$

The power series in D corresponding to $I_{|f|}$ will be of the form $G(D) = A_0 + A_1D + \ldots$, and $A_0 \neq 0$ (by Case IV of Theorem 8). Thus it follows from Theorem 11 that

$$I_{|f|\Lambda}1 = -A_0^{-1}A_1 + O(\log^{1-\delta}x).$$

Thus, since $h(x) = O(1)$

$$\begin{aligned} |h(x)|\log x &< I_{|f|\Lambda}|h(x)| + O(\log^{2-\delta}x) + O(1) \\ &< O(1).I_{|f|\Lambda}1 + O(\log^{2-\delta}x) + O(1) = O(\log^{2-\delta}x) + O(1). \end{aligned}$$

Consequently $|h(x)| < O(\log^{1-\delta}x) + O(\log^{-1}x)$. That is,

THEOREM 12*. *If f is a character satisfying*

$$\sum_{Nw<x} f(w) = O(x/\log^{1+\delta}x)$$

and

$$\sum_{Nw<x} |f(w)| = O(x/\log^{1+\delta} x),$$

then

$$\sum_{Nw<x} f(w)\,\Lambda(w) = O(x/\log^{1-\delta} x) + O(x/\log x).$$

8. Proofs of Theorems A and B. We return now to the situation of §2. Let H denote a generic class W/W' and let $e_H(w)$ be the characteristic function of H, that is, $e_H(w) = 1$ if $w \in H$ and zero otherwise. From the properties of characters (2.2) and (2.3) we have the relations

$$\chi = \sum_H \chi(H)e_H, \qquad e_H = h^{-1}\sum_\chi \bar{\chi}(H)\chi. \tag{8.1}$$

We assumed in (2.4) that

$$S_{e_H}1 = \sum_{\substack{Nw<x\\ w\epsilon H}} 1 = c_H x + O(x/\log^\gamma x), \qquad \sum c_H > 0. \tag{8.2}$$

Thus

$$S_\chi 1 = \sum_H \chi(H)S_{e_H}1 = A_\chi x + O(x/\log^\gamma x), \qquad A_\chi = \sum_H \chi(H)c_H,$$

and for the identity $\chi_0 = E$. Also

$$S_{\chi_0}1 = cx + O(x/\log^\gamma x), \qquad A_{\chi_0} = c = \sum c_H > 0.$$

The characters are thus functions of the type which were dealt with in the preceding sections. Let

$$L_\chi(D) = \sum_{\nu=-1}^{\infty} L_\nu(\chi)D^\nu; \qquad L_{-1}(\chi) = A_\chi$$

be the polynomial corresponding to I_χ in (6.6), then we distribute the characters of K in three classes

$$\Gamma_1 = \{\chi; \chi \in K, A_\chi = L_{-1}(\chi) \neq 0\},$$
$$\Gamma_2 = \{\chi; \chi \in K, A_\chi = 0, L_0(\chi) \neq 0\},$$
$$\Gamma_3 = \{\chi; \chi \in K, A_\chi = 0, L_0(\chi) = 0\}.$$

Theorem 8 now implies that (all characters are in our case subjected to CasesI–III):

COROLLARY 3.

$$I_{\chi n} = L_\chi^{-1}(D) + O(1) + \begin{cases} O(\log^{n-\delta} x), & \chi \in \Gamma_1 \\ O(\log^{n+1-\delta} x), & \chi \in \Gamma_2 \\ O(\log^{n+2-\delta} x), & \chi \in \Gamma_3. \end{cases}$$

From Theorem 12 we now obtain Theorem A.

Theorem A and (8.1) yield

$$\sum_{Nw<x} e_H(w)\,\Lambda(w) = \sum_{\substack{Nw<x\\ w\in H}} \Lambda(w) = h^{-1}\sum_{Nw<x} \chi(H)\,\bar{\chi}(w)\,\Lambda(w)$$

$$= h^{-1}\Big[\sum_{\chi\in\Gamma_1} \chi(H) - \sum_{\chi\in\Gamma_2} \chi(H)\Big]x + o(x) = d_H x + o(x).$$

From here we can follow the ideas developed in **(4)**, but replacing the "Dirichlet density" k of a set S, defined there, by the sum

$$\sum_{w\in S} d(w)^{-1}\Lambda(w) = k\log x + O(1),$$

by dealing in a parallel way with the sum

$$\sum_{w\in S} \Lambda(w) = kx + o(x)$$

and by calling k the Dirichlet density of the set S. As the reasoning is identical with that of **(4**, p. 602**)** as well as the passage from

$$\psi_H(x) = \sum_{\substack{Nw<x\\ w\in H}} \Lambda(w)$$

to

$$\pi_H(x) = \sum_{\substack{Np<x\\ p\in H}} 1$$

we just quote the final results.

(a) Γ_1 is a subgroup of index 1 or 2 in the group $\Gamma_1 \cup \Gamma_2$.

(b) Let $U = \{w; \chi(w) = 1, \chi \in \Gamma_1\}$, then $W' \subseteq U \subseteq W$ and U/W' has Γ_1 as the group of characters. Put $U^* = \{w; w \in U, \chi(w) = 1 \text{ for all } \chi \in \Gamma_2\}$ then $U \supseteq U^* \supseteq W'$ and Theorem B is valid for these groups U and U^*. (Compare with **(4**, Theorem 3.1**)**.)

9. Other "elementary results." In the present section we shall outline the extensions of the elementary proofs of **(1)** to our case. These lie in the following extension of **(1**, Theorem 9.2**)**.

THEOREM 13. *Let* $g \in C(W)$; $G(D) = \gamma_{-q}D^{-q} + \gamma_{-q+1}D^{-q+1} + \dots$. *Let f be a character satisfying* (6.1) *and* (6.2). *Then for* $n < \delta - q - 1$

$$I_{f^{-1}*g}\log^n x = [F^{-1}(D)G(D)]\log^n x + o(1),$$

where $F(D)$ is given in (6.6), *provided that the following conditions hold:*

(i) $I_{f^{-1}}R_\nu(x; g, G) = O(1)$ for $\nu < n + 1$,

(ii) $I_{|f|}\log x\, R_n(x; g, G) = o(\log x)$

(iii) $I_{|h|}1 = O(x^\theta)$ where $h = f^{-1}*g$ and $\theta < 1$

(iv) $\sum_{x<Nw<tx} (Nw)^{-1}|h(w)| = o(1)$ as $(t, x) \to (1, \infty)$

(v) Case I and $q < 2$; Case II and $q < 1$; Case III and $q < 0$.

Before proceeding with this proof we give here some examples as applications.

Example 1. $g = \epsilon$ is the identity, and $G(D) = 1$. Thus $R_\nu(x; \epsilon, 1) = O(1)$ for all ν which yields (i) and (ii) trivially. Here, $h = f^{-1}*\epsilon = f^{-1}$ hence $I_{|h|}1 = O(\log x)$, by (6.3) and (6.2) which proves (iv). To prove (v) we obtain by (6.3) and (6.9) that for some $K > 0$,

$$0 < \sum_{x<Nw<tx} (Nw)^{-1}|f^{-1}(w)| < K \sum_{x<Nw<tx} (Nw)^{-1}|f(w)| = KA \log t + O(\log^{-\delta} x)$$
$$= o(1)$$

as $(t, x) \to (1, \infty)$. Consequently,

COROLLARY 4.

$$I_{f^{-1}} \log^n x = F^{-1}(D) \log^n x + o(1) \quad \text{for} \quad n < \delta - 1.$$

In particular this yields, for $\delta > 1$:

$$I_{f^{-1}}1 = \sum_{Nw<x} \frac{\mu(w)f(w)}{Nw} = \begin{cases} o(1) & \text{in Case I} \\ \alpha_0^{-1} + o(1) & \text{in Case II} \\ \alpha_1^{-1} \log x + \alpha_2 \alpha_1^{-1} + o(1) & \text{in Case III.} \end{cases}$$

This includes for the case $f(w) = 1$, one of the equivalent forms of the prime number theorem, but our effort to follow the classical proof of that theorem from this result failed, and we have obtained the prime number theorem in §8 in a different way.

Example 2. $g(w) = (fL)(w) = f(w) \log Nw$, and $G(D) = -F'(D)$. We shall consider only the case $A \neq 0$ (in Case I, $q = 2$, Cases II and III, $q < 0$). In this example $R_n(x; fL, -F') = O/\log^{n+1-\delta} x$ (Theorem 7) and thus Corollary 2 implies that

$$I_{f^{-1}} R_\nu(x; fL, -F') = O(\log^{\nu+2-\delta} x) + O(1) = O(1)$$

for all ν satisfying $\nu + 2 - \delta < 0$. This proves the validity of (i) for all $n + 3 - \delta < 0$. But then (ii) also holds since

$$I_f \log x \, R_n(x; fL, -F') = I_f O(\log^{n+2-\delta} x) = O(1) + O(\log^{n+3-\delta} x) = O(1).$$

Now $h = f^{-1}*fL = f\Lambda$ which implies the validity of (iii) by Theorem 9 when applied to $|f|$, since $I_{|f|}1 = O(\log x) + O(\log^{2-\delta} x)$. The last condition (iv) follows readily from (7.3). Consequently we obtain by applying Theorem 13 that

COROLLARY 5.

$$I_{f\Lambda} \log^n x = -[F^{-1}(D)F'(D)] \log^n x + o(1) \text{ if } \delta > n + 3.$$

In particular, if $\sigma = +1, 0, -1$ in Cases I, II, III respectively, then for $\delta > 3$:

$$\sum_{Nw \leq x} \frac{f(w)\,\Lambda(w)}{Nw} = \sigma \log x + \beta + o(1)$$

which is another equivalent form of the prime number theorem in the case of integers with $f(w) = 1$. Note that this is obtained only for $\delta > 3$ (that is, $\gamma > 4$) whereas the other equivalent for

$$\sum_{Nw \leq x} f(w)\,\Lambda(w) = \sigma x + o(x)$$

was obtained in Theorem 12 for $\gamma > 2$.

Now for the proof of Theorem 13. We wish to show that the function $h(x) = R_n(x; f^{-1}*g, F^{-1}G)$ satisfies the requirements of Theorem 10 with $g(w) = |f(w)|\,\Lambda_2(w)$. It was already proved in the preceding that $g(w)$ satisfies (g1) of Theorem 10 if $\delta > 1$, and we now prove the validity of (g2). It follows from Theorem 3 that

$$R_n(x; g, G) = R_n(x; f*(f^{-1}*g),$$
$$F(F^{-1}G)) = I_f R_n(x; f^{-1}*g, F^{-1}G) + \sum_{j=0}^{n+p} c_j R_j(x; f, F) + a,$$

where D^{-p} is the first power of D in $F^{-1}G$, and for some constants $c_j\, a$ ($a = 0$ if f is a character satisfying Cases II and III). Operating with $I_{f^{-1}}\log x$ on this result, we find

$$\begin{aligned} &I_{f^{-1}} \log x\, I_f R_n(x; f^{-1}*g, F^{-1}G) \\ &\quad = I_{f^{-1}} \log x R_n(x; g, G) - \sum_{j=0}^{n+p} c_j\, I_{f^{-1}} \log x R_j(x; f, F) - a I_{f^{-1}} \log x \\ &\quad = o(\log x) + O(\log^{n+p+1-\delta} x) + O(1) + O(\log^{1-\delta} x) \\ &\quad = o(\log x) \end{aligned}$$

if $n + p + 1 - \delta < 1$. Since $R_j(x; f, F) = O(\log^{j-\delta} x)$ by Theorem 6 the rest follows like the proof of Theorem 12, that is

$$I_{f^{-1}} \log x\, I_f\, R_n(x; f^{-1}*g) = (\log x + I_{f\Lambda}) R_n(x; f^{-1}*g) = o(\log x),$$

from which we deduce that

$$|R_n(x; f^{-1}*g)| \log^2 x \leq I_{|f|\Lambda_2} |R_n(x; f^{-1}*g)| + o(\log^2 x),$$

namely, (g2).

To prove conditions (h1) – (h3), we first observe that

$$R_n(f^{-1}*g) = I_{f^{-1}} R_n(g) + \sum_{j=0}^{n+q} c_j R_j(f^{-1}) + a = O(1) + O(\log^{n+q-\delta} x)$$

where $a \neq 0$ if the situation is of Case III. This shows that

$$R_n(f^{-1}*g) = O(1) \text{ and } R_{n+1}(f^{-1}*g) = O(1)$$

if $n + 1 + q - \delta < 0$.

The completion of the proof follows the computation of **(1**, pp. 306–7**)**. We do not repeat the computation but present the final result in the following proposition.

PROPOSITION 9. *Let* $l(w) \in C(W)$ *and*

$$L(D) = \sum_{\nu=-m}^{\infty} \gamma_\nu D^\nu.$$

Then

$$(9.1)\qquad \sum_{\mu<x} \mu^{-1}R_n(\mu; l, L) = (n+1)^{-1}R_{n+1}(x; l, L) + O(1) + O\Big(\sum_{Nw<x} (Nw)^{-2}|l(w)| \log^n Nw\Big)$$

and

$$(9.2)\qquad R_n(tx; l, L) - R_n(x; l, L) = \sum_{Nw<x} (Nw)^{-1}l(w) \log^n(tx/Nw) + \sum_{j=0}^{n-1} \binom{n}{j} \log^{n-j} tR_j(x; l, L) + O(\log^{m-1} x \log^{n+1} t),$$

woth the last factor omitted if $L(D)$ *does not contain negative powers of* D.

Thus in our case, $l = f^{-1}*g$, $L = F^{-1}G$ we observe that (iii) of Theorem 13 yields, by Lemma 2:

$$\sum_{Nw<x} (Nw)^{-2}|l(w)| \log^n Nw = O(x^\delta)x^{-1} \log^n x - \int_1^{x} O(t^\delta) d[t^{-1}\log^n t] = O(1)$$

which shows that $R_{n+1}(f^{-1}*g) = O(1)$ implies (h2). Condition (h3) of Theorem 10 follows from (iv), since in our case $R_j(x;l,L) = O(1)$ and $m = 1$ (in Case III, we have to require that $q = 0$) imply that

$$\sum_{x<Nw<tx} (Nw)^{-1}l(w)\log^n[tx/Nw] = o(1).$$

This completes the proof that if $n + q + 1 < \delta$ and (a) Case I with $q < 2$, (b) Case II, $q < 1$, or (c) Case III, $q < 0$, all conditions of Theorem 10 are fulfilled. Thus the proof of Theorem 13 is complete.

10. The character $f(n) = n^{it}$. We conclude our result with an application for the semi-group of integers and the character $f(n) = n^{it}$, $t \neq 0$ fixed.

Clearly f is a character, and satisfies $I_{|f|}1 = \sum n^{-1} = \log x + c + O(x^{-1})$ and $I_f 1 = \sum n^{-1+it} = c_t + O(x^{-1})$ where $c_t = \zeta(1 - it)$. Thus this function f satisfies the condition of either Case II or Case III.

If (III) is valid then we would get from Theorem 13 (Example 2) that

$$I_{f^{-1}*fL}1 = I_{f\Lambda}1 = \sum_{n<x} \Lambda(n)n^{-1+it} = -\log x + c + o(1).$$

But since

$$\sum_{n\leq x}\frac{\Lambda(n)}{n} = \log x + c_0 + o(1)$$

(the case $f(n) = 1$) we have by Lemma 2

$$-\log x + c + o(1) = \sum_{n\leq x}\Lambda(n)n^{-1+it}$$

$$= [\log x + c_0 + o(1)]x^{it} - \int_1^x [\log u + c_0 + o(1)]du^{it}$$

$$= O(1) + \int_1^x o(1)du^{it} = o(\log x).$$

Indeed, if the function $|o(1)| < \epsilon$ for $x > \lambda$ and $|o(1)| < K$ for $x \leq \lambda$, then

$$\left|\int_1^x o(1)du^{it}\right| < Kt\int_1^\lambda u^{-1}du + t\epsilon\int_\lambda^x u^{-1}du < M + \epsilon t\log x. \qquad \text{Q.E.D.}$$

Thus Case III is disposed of and there remains Case II, which means that

$$\sum_{n=1}^{\infty} n^{-1+it} = \zeta(1-it) \neq 0$$

and consequently that the series

$$\sum_{n=1}^{\infty} n^{-1+it}\Lambda(n)$$

converges for all $t \neq 0$. From this one readily proves that

$$\sum_p p^{-1+it}\log p$$

converges.

References

1. S. A. Amitsur, *On arithmetic functions*, J. d'Anal. Math., *5* (1956–7), 273–317.
2. ——— *Some results on arithmetic functions*, J. Math. Soc. Japan *11* (1959), 275–290.
3. A. Beurling, *Analyse de la loi asymptotique de la distribution des nombre premiers generalises*, Acta. Math., *68* (1937), 255–291.
4. H. Shapiro and W. Forman, *Abstract prime number theorems*, Comm. Pure and Applied Math., *7* (1954), 587–619.
5. K. Yamamoto, *Theory of arithmetic linear transformations and its applications to an elementary proof of Dirichlet's theorem*, J. Math. Soc. Japan, 7 (1955), 424–434.
6. ——— *Arithmetic linear transformations in an algebraic number field*, Mem. Fac. Sci., Kyushu Univ. Sci. A. *12* (1958), 41–66.
7. Hardy and Wright, *The theory of numbers* (Oxford, 1954).

Hebrew University

BLOCK DESIGN GAMES

A. J. HOFFMAN AND MOSES RICHARDSON[1]

In this paper, we define and begin the study of an extensive family of simple n-person games based in a natural way on block designs, and hitherto for the most part unexplored except for the finite projective games **(13)**. They should serve at least as a proving ground for conjectures about simple games. It is shown that many of these games are not strong and that many do not possess main simple solutions. In other cases, it is shown that they have no equitable main simple solution, that is, one in which the main simple vector has equal components. On the other hand, the even-dimensional finite projective games $PG(2s, p^n)$ with $s > 1$ possess equitable main simple solutions, although they are not strong either. These results are obtained by means of the study of the possible blocking coalitions. Interpretations in terms of graph theory, network flows, and linear programming are discusssed, as well as k-stability, automorphism groups, and some unsolved problems.

1. Preliminaries on block designs. Block designs have long been studied from various points of view and have an extensive literature, an introduction and references to which can be found in Hall **(8)**.

By a *block design*[2] we shall mean a set N of v elements $\{1, 2, \ldots, v\}$, and a family of b distinguished subsets $W_1, W_2, \ldots, W_b$ of N called *blocks*, such that

(a) every W_i contains k elements, $k < v$,

(b) every element x belongs to r blocks.

A block design may be specified by means of its *incidence matrix* $A = ||a_{ij}||$ with v rows and b columns, where $a_{ij} = 1$ if the ith element belongs to the jth block and $a_{ij} = 0$ if not. The numbers v, b, k, r are termed *parameters* of the design. Clearly,

$$vr = bk, \tag{1}$$

since each side represents the total number of ones in the incidence matrix. A block design is termed *symmetric* if $v = b$ or, equivalently, $k = r$. A block design is termed *balanced* if every two elements occur together in λ blocks. The numbers v, b, k, r, λ are termed the *parameters* of the balanced block design and satisfy, in addition to (1), the relation

$$r(k - 1) = \lambda(v - 1). \tag{2}$$

[1]Received January 4, 1960. Some of the work of this paper was done while this author was partly supported by a National Science Foundation Faculty Fellowship.

[2]Also referred to as incomplete block design and tactical configuration in the literature. Cf. (**2**; **3**; **4**; **10**; **11**).

A symmetric balanced block design is often referred to as a (v, k, λ)-system. Perhaps the most familiar balanced block designs are the finite geometries, projective and euclidean, where the points are taken as the elements and the lines as the blocks; for these, we have $\lambda = 1$ since two points determine a line. Other balanced block designs which have long been studied are the *Steiner triple systems* (cf. **8**; **10**; **11**) for which either $v = 6t + 1$, $b = t(6t + 1)$, $r = 3t$, $k = 3$, $\lambda = 1$, or $v = 6t + 3$, $b = (2t + 1)(3t + 1)$, $r = 3t + 1$, $k = 3$, $\lambda = 1$. The Steiner triple systems with $v = 1, 3, 7$ we shall here term *trivial*. The case $v = 7$ is the familiar seven-point projective plane.

A block design is termed *partially balanced* if:

(A) There exist non-negative integers $\lambda_1, \lambda_2, \ldots, \lambda_h$ and positive integers $n_1, n_2, \ldots, n_h$ such that to every element x corresponds n_j other elements, called jth *associates* of x, with the property that any jth associate of x occurs together with x in λ_j blocks, and

(B) if x and y are ith associates then the number of elements which are jth associates of x and kth associates of y is $p_{jk}{}^i$.

The numbers $v, b, k, r, \lambda_1, \ldots, \lambda_h, n_1, \ldots, n_h, p_{jk}{}^i$ are termed the *parameters* of the partially balanced block design. It is understood that the numbers $\lambda_i, n_i, p_{jk}{}^i$ are independent of the choice of element. We shall suppose that $h > 1$. If $h = 1$, so that all λ_i may be replaced by λ and all n_i by $v - 1$, then the block design is balanced.

A partially balanced block design with two associate classes is termed *group divisible* if the elements can be divided into m groups each with n elements so that pairs of elements in the same group occur together in λ_1 blocks and pairs of elements in different groups occur together in λ_2 blocks, $\lambda_1 \neq \lambda_2$. It is clear that $n_1 = n - 1$ and $n_2 = n(m - 1)$. A group divisible design (cf. **2**; **5**) is termed *singular* if $r = \lambda_1$, *semi-regular* if $r > \lambda_1$ and $rk = v\lambda_2$, *regular* if $r > \lambda_1$ and $rk > v\lambda_2$.

Let s_{ij} be the number of elements common to the ith and jth blocks of a design; the matrix $S = \|s_{ij}\| = A^T A$. It is known that in a symmetric balanced block design all $s_{ij} = \lambda$. It is also known (cf. **5**) that: for a regular symmetric group divisible design,

$$\begin{aligned} &\lambda_2(r - \lambda_1)/(r^2 - v\lambda_2) \leqslant s_{ij} \leqslant \lambda_1 \quad \text{if} \quad \lambda_1 > \lambda_2, \\ &\lambda_1 \leqslant s_{ij} \leqslant \lambda_2(r - \lambda_1)/(r^2 - v\lambda_2) \quad \text{if} \quad \lambda_1 < \lambda_2; \end{aligned} \tag{3}$$

for a symmetric regular group divisible design with $r^2 - v\lambda_2$ and $\lambda_1 - \lambda_2$ relatively prime, all

$$s_{ij} = \lambda_1 \quad \text{or} \quad \lambda_2; \tag{4}$$

for a symmetric semi-regular group divisible design

$$\lambda_1 \leqslant s_{ij} \leqslant \frac{2\lambda_2 r^2}{r + v\lambda_2 - \lambda_1} - \lambda_1. \tag{5}$$

If the block design D^* has parameters v^*, b^*, k^*, r^* and incidence matrix

A, then the *dual* block design D has parameters $v = b^*, b = v^*, k = r^*, r = k^*$ and incidence matrix A^T, the transpose of A (cf. **3**; **15**).

Restriction. We shall henceforth confine ourselves to block designs, designated by D, of which the incidence matrices have no two column vectors equal, and correspondingly to block designs, designated by D^*, of which the incidence matrices have no two row vectors equal. Thus no two blocks of a design D are to be equal sets.[3]

If D^* is a partially balanced design with all $\lambda^*_i > 0$, then in the dual design D every pair of distinct blocks has a non-empty intersection. If D^* is a balanced design with $\lambda^* > 0$, then in the dual design D the intersection of every pair of distinct blocks has λ^* elements.

2. Preliminaries on simple games. Let N be a finite set $\{1, 2, \ldots, v\}$ of v *players*. Let $\mathfrak{N}$ be the class of all subsets of N, each of which is called a *coalition*. If $\mathfrak{S} \subset \mathfrak{N}$, let $\mathfrak{S}^+$ be the class of all supersets of elements of $\mathfrak{S}$, and let $\mathfrak{S}^*$ be the class of all complements of elements of $\mathfrak{S}$. By a *simple game* is meant an ordered pair $G = (N, \mathfrak{W})$ where $\mathfrak{W}$ is a subclass of $\mathfrak{N}$ satisfying

$$(\alpha) \qquad \mathfrak{W} = \mathfrak{W}^+$$

$$(\beta) \qquad \mathfrak{W} \cap \mathfrak{W}^* = \phi.$$

Elements of $\mathfrak{W}$ are termed *winning coalitions*, elements of $\mathfrak{L} = \mathfrak{N} - \mathfrak{W}$ are termed *losing coalitions*, and elements of $\mathfrak{B} = \mathfrak{L} \cap \mathfrak{L}^*$ are termed *blocking* coalitions. The simple game[4] G is termed *strong* if and only if $\mathfrak{B} = \phi$. A simple game may be defined by specifying the class $\mathfrak{W}^m \subset \mathfrak{W}$ of *minimal winning coalitions*, by virtue of condition (α). Let $W_1, W_2, \ldots, W_b$ be the minimal winning coalitions.

A *dummy* is a player i such that $f(S \cup \{i\}) = f(S)$ for all $S \in \mathfrak{N}$ where f is the characteristic function of the game. We shall confine ourselves here to *strictly essential* games, that is, having no dummies. We use the $0 - 1$ normalization.

A vector $(a_1, a_2, \ldots, a_v)$ of non-negative real components such that

$$(6) \qquad \sum_{i \in S} a_i = 1 \quad \text{for} \quad S \in \mathfrak{W}^m,$$

$$(7) \qquad \sum_{i \in S} a_i > 1 \quad \text{for} \quad S \in (\mathfrak{W} \cup \mathfrak{B}) - \mathfrak{W}^m$$

is termed a *main simple vector* (cf. **12**; **14**; **7**). If there exists a main simple vector, then the finite set of imputations $X = \{x^{(S)} | S \in \mathfrak{W}^m\}$ where

$$x_i^{(S)} = \begin{cases} a_i & \text{if} \quad i \in S \\ 0 & \text{if} \quad i \notin S \end{cases}$$

[3]It is possible for a design with distinct blocks to have a dual with two or more blocks equal; see for example, design S12 in **(2)**. If, in a balanced design, $\lambda < r$, then two row vectors of the incidence matrix cannot be equal. If, in a partially balanced design, all $\lambda_i < r$, the same conclusion holds.

[4]In **(12)**, the terminology is such that all simple games are those which are termed strong here and in **(14)**. We shall use the terminology of **(14)** throughout.

is termed a ***main simple solution*** of the simple game G. A player i is ***indifferent*** relative to the main simple solution X if $a_i = 0$. This can occur without i necessarily being a dummy. ***We shall suppose that there are no indifferent players throughout.*** A main simple solution will be termed *equitable* if all the components a_i of its main simple vector are equal.

A necessary condition for a set $X \subset N$ to be a blocking coalition is that the row vectors R_i, $i \in X$, in the incidence matrix, with columns corresponding to minimal winning coalitions and rows corresponding to players, shall have a boolean sum equal to the unit vector $\mathfrak{U}_b$, all b components of which are ones; that is,

$$\sum_{i \in X} R_i = \mathfrak{U}_b$$

where the summation is boolean.

3. Block design games. Any block design D, subject to the restriction at the end of § 1, in which, furthermore, every pair of blocks has a non-empty intersection, may be used to define a simple game, called a ***block design game***, in which the players correspond to the elements of the design and the minimal winning coalitions correspond to the blocks of the design. Particular examples are the finite projective games studied in **(13)**, symmetric balanced block designs, group divisible designs with all $s_{ij} > 0$, the duals of finite euclidean planes, the duals of Steiner triple systems, the duals of balanced block designs with $\lambda > 0$, and the duals of partially balanced block designs with all $\lambda_i > 0$.

The following lemmas will be useful.

LEMMA 1. *If, in any simple game, there exists a blocking coalition B (properly) contained in some minimal winning coalition W, then there exists no main simple solution.*

Proof. If there were a main simple vector we would have

$$\sum_{i \in W} a_i = 1 \qquad \text{but} \qquad \sum_{i \in B} a_i \not> 1.$$

LEMMA 2. *If every blocking coalition B in a block design game is such that the number of players in B is greater than k, then there exists an equitable main simple solution.*

Proof. We can take $a_i = 1/k$.

LEMMA 3. *If, in a block design game, there exists a blocking coalition B of which the number of players is less than or equal to k, then there exists no equitable main simple solution.*

Proof. For we would have

$$\sum_{i \in B} a_i \leqslant 1$$

and therefore not > 1 as required.

4. Some theorems on block design games. We establish some theorems concerning blocking coalitions and main simple solutions of various block design games. Examples are collected in § 8.

THEOREM 1. *A block design game is not strong if one of the following conditions hold:*
(a) $v = 2k$, $b < \frac{1}{2}C(v, k)$,
(b) $v < 2k$, $b < C(v, k)$,
(c) $v > 2k$, *and some* $(v - k + 1)$*-tuple of players constitutes a losing coalition.*

Proof. Under hypothesis (a), at least one k-tuple of players is not in $\mathfrak{W}^m$ and has its complementary k-tuple also not in $\mathfrak{W}^m$, for the number of k-tuples in $\mathfrak{W}^m \cup \mathfrak{W}^{m*}$ is $2b < C(v, k)$. Hence $\mathfrak{B} = \mathfrak{L} \cap \mathfrak{L}^* \neq \phi$.

Under hypothesis (b), there exists a k-tuple not in $\mathfrak{W}^m$ whose complementary $(v - k)$-tuple is not in $\mathfrak{W}$ since $v - k < k$, while any set in $\mathfrak{W}$ has at least k members.

Under hypothesis (c), the complement of the given $(v - k + 1)$-tuple is also in $\mathfrak{L}$ since it has only $k - 1$ members.

More precise information concerning blocking coalitions in various block design games is given in the remaining theorems of this section.

THEOREM 2. *In any simple game, if there exists a player x_1 and a minimal winning coalition W_1 containing x_1 such that every other minimal winning coalition W containing x_1 intersects W_1 in more than one element, then there exists a blocking coalition B which is a (proper) subset of W_1. Hence, under these hypotheses, the game is not strong and there exists no main simple solution.*

Proof. Let $W_1 = \{x_1, x_2, \ldots, x_k\}$, say, and let $B = W_1 - \{x_1\} = \{x_2, \ldots, x_k\}$. Now every minimal winning coalition W different from W_1 must intersect W_1, and furthermore, by hypothesis, must intersect B. Consequently, B is a blocking coalition (properly) contained in W_1. The last sentence of the theorem follows from Lemma 1.

COROLLARY. *The hypotheses and hence the conclusions of Theorem 2 are satisfied if the block design game D is any of the following:*

(a) *a symmetric balanced block design with $\lambda > 1$;*

(b) *the dual of any balanced block design with $\lambda > 1$; in particular, the dual of the design formed by the s-spaces in a projective or euclidean m-space $PG(m, p^n)$ or $EG(m, p^n)$ with $1 < s < m$;*

(c) *the dual of a partially balanced block design with all $\lambda_i > 1$;*

(d) *a symmetric regular group divisible design with $1 < \lambda_1 < \lambda_2$;*

(e) *a symmetric regular group divisible design with $\lambda_1 > \lambda_2$ and $\lambda_2(r - \lambda_1) > r^2 - v\lambda_2$;*

(f) *a symmetric semi-regular group divisible design with $\lambda_1 > 1$;*

(g) *a symmetric regular group divisible design with $r^2 - v\lambda_2$ and $\lambda_1 - \lambda_2$ relatively prime and both λ_1 and λ_2 greater than one.*

THEOREM 3. *If D^* is a balanced block design with parameters v^*, b^*, k^*, r^*, and $\lambda^* = 1$, then its dual D yields a game which is not strong if either*[5] *$k = 3$ and $r \geqslant 4$, or $k \geqslant 4$ and $r \geqslant 3$. In particular, under these hypotheses, there exists a blocking coalition with $k + r - 2$ members.*

Proof. In D we have $bk = vr$ and $k(r - 1) = b - 1$, and the intersection of every pair of minimal winning coalitions has just one element. Consider any block $W = \{x_1, \ldots, x_k\}$ of D. There are

$$r + (k - 2)(r - 1) = k(r - 1) - r + 2 = b - (r - 1)$$

blocks containing at least one of the elements $x_1, \ldots, x_{k-1}$, leaving $r - 1$ blocks intersecting W in x_k only. In these $r - 1$ blocks there are $(r - 1)(k - 1)$ elements other than x_k. There exist $(k - 1)^{r-1}$ possible $(r - 1)$-tuples with one element chosen from each of these $r - 1$ blocks. Excluding W, there are $(r - 1)(k - 1)$ blocks containing elements of $\{x_1, \ldots, x_{k-1}\}$. But, if $k = 3$ and $r \geqslant 4$, or if $k \geqslant 4$ and $r \geqslant 3$, then

$$(k - 1)^{r-1} > (k - 1)(r - 1).$$

Therefore, in this case, at least one such $(r - 1)$-tuple $\{y_1, \ldots, y_{r-1}\}$ exists not forming a block together with any member of $\{x_1, \ldots, x_{k-1}\}$. Thus, $\{x_1, \ldots, x_{k-1}, y_1, \ldots, y_{r-1}\}$ is a blocking coalition, which completes the proof.

COROLLARY 1. *The dual D of a non-trivial Steiner triple system D^* is not strong.*

Proof. Except for the cases $v^* = 1, 3, 7$, which we have termed trivial, the Steiner triple systems have $k^* = 3$ and $r^* \geqslant 4$. Hence, in the dual, $r = 3$ and $k \geqslant 4$.

COROLLARY 2. *If D^* is the system of lines in the finite euclidean space $EG(m, p^n)$ of m dimensions over the Galois field $GF(p^n)$ for $m \geqslant 2$ and $p^n \geqslant 3$, then the dual D is not strong.*

Proof. In D^* we have $v^* = p^{nm}$, $b^* = p^{n(m-1)}(1 + p^n + \ldots + p^{n(m-1)})$, $k^* = p^n$, $r^* = 1 + p^n + \ldots + p^{n(m-1)}$, and $\lambda^* = 1$. Hence in the dual, $r \geqslant 3$ and $k \geqslant 4$.

COROLLARY 3. *If D^* is the system of lines in the finite projective space $PG(m, p^n)$ of m dimensions over the Galois field $GF(p^n)$ with $m \geqslant 3$ and $p^n \geqslant 2$, then the dual D is not strong.*

Proof. In D^*, we have $v^* = 1 + p^n + \ldots + p^{mn}$,

$$b^* = \frac{(1 + p^n + \ldots + p^{mn})(1 + p^n + \ldots + p^{(m-1)n})}{1 + p^n}, \; k^* = 1 + p^n,$$

[5]It is easily verified that the only remaining cases are the triangle and the seven-point projective plane, which are strong, and the duals of complete n-gons, $n \geqslant 4$, also termed triangular association schemes below, which are not strong.

$r^* = 1 + p^n + \ldots + p^{(m-1)n}$, and $\lambda^* = 1$. Hence, in D, we have $k > 4$, and $r > 3$.

5. Some games with no equitable main simple solution.

THEOREM 4. *If D^* is the system of hyperplanes in the finite euclidean m-space $EG(m, p^n)$, $m > 2$, then in the dual D there exists a blocking coalition with p^n members.*

Proof. In $EG(m, p^n)$, consider any family of p^n parallel hyperplanes or $(m-1)$-spaces, one through each point of a transversal line. Their union contains all the points of the space. In the dual D, these hyperplanes correspond to p^n elements incident with all the blocks, no two of which elements occur together in any block. Therefore, these elements constitute a blocking coalition with p^n members.

COROLLARY. *The block design game D, dual to the system of hyperplanes of a finite euclidean m-space $EG(m, p^n)$, $m > 2$, is not strong and has no equitable main simple solution.*

Proof. The last conclusion follows at once from Lemma 3 of § 3.

THEOREM 5. *In the dual D of a non-trivial Steiner triple system D^*, there exists a blocking coalition with k members.*

Proof. In D^*, consider the set of r^* triples containing a given element x, say. Delete any two of these triples, say (x, a, b) and (x, c, d) where a, b, c, and d are, of course, distinct elements. Then there exist triples (a, c, y) and (b, d, z) with $y \neq x$, $z \neq x$ in D^*, since $\lambda^* = 1$ and since the trivial systems have been excluded. Replacing the two deleted triples by the latter two, we have a set of triples whose union contains all elements of D^* and does not contain any set of all triples through any particular element. In the dual D, this corresponds to a set of elements incident with all blocks but not containing all elements of any particular block. This is a blocking coalition with $k = r^*$ elements.

COROLLARY. *The block design game D, dual to a non-trivial Steiner triple system D^*, is not strong and has no equitable main simple solution.*

Proof. The latter conclusion follows at once from Lemma 3 of § 3.

Remark 1. The system of all lines in m-dimensional finite projective space $PG(m, 2)$ over the integers modulo 2, and the system of all lines in m-dimensional finite euclidean space $EG(m, 3)$ over the integers modulo 3, are Steiner triple systems. Of course, not every Steiner triple system is of this type.

Remark 2. The conclusion of Theorem 3 does not hold for the dual of a partially balanced design with some $\lambda_i = 1$ and some $\lambda_j > 1$. For instance, the game of Example 2 is not strong but the game of Example 6 is strong (see § 10, below).

By a ***triangular association scheme*** (cf. **2**) is meant an n by n matrix in which: (a) the elements on the principal diagonal are left blank; (b) the $n(n-1)/2$ positions above the principal diagonal are filled by the numbers $1, 2, \ldots, n(n-1)/2$; (c) the matrix is symmetric. If we take the players to be the numbers $1, 2, \ldots, n(n-1)/2$, and the minimal winning coalitions $W_1, \ldots, W_n$ to be the rows of the triangular association scheme, then it is easily seen that we have a block design game with $v = n(n-1)/2$, $b = n$, $k = n-1$, $r = 2$, and every two distinct minimal winning coalitions have one player in common. We shall term such a game a ***triangular game***.

THEOREM 6. *A triangular game with $n > 3$ has a blocking coalition with $\leqslant k-1$ members; hence it is not strong and has no equitable main simple solution.*

Proof. Let $\{x_1\} = W_1 \cap W_2$. Since $W_1 \cup W_2$ has $2n-3$ members, and $v > 2n-3$ for $n > 3$, there exists an $x_2 \notin W_1 \cup W_2$. Suppose, for example, that $\{x_2\} = W_3 \cap W_4$. Then we can choose arbitrarily $x_i \in W_i$ $(i = 5, 6, \ldots, n)$, distinct or not. Obviously, the distinct members of the set $\{x_1, x_2, x_5, x_6, \ldots, x_n\}$ form a blocking coalition with $\leqslant n-2 = k-1$ members. The second assertion of the theorem follows from Lemma 3.

Remark 3. In fact, it is easy to see that there exists a blocking coalition with $[(n+1)/2]$ members, where $[x]$ is the largest integer $\leqslant x$, but this stronger result does not seem to have any interesting game-theoretic implications.

6. Even-dimensional finite projective games. In **(13)**, finite projective games $PG(h, p^n)$ were defined as follows: the players are the points of the finite projective space $PG(h, p^n)$ of dimension $h > 1$ over the Galois field $GF(p^n)$, and the minimal winning coalitions are the $(s+1)$-spaces if $h = 2s + 1$, and the s-spaces if $h = 2s$. As noted in **(13)**, the odd-dimensional finite projective games are not strong and have no main simple solution since the s-spaces are blocking coalitions contained in the minimal winning coalitions (cf. Lemma 1, above). In **(13)**, it is also proved that the plane games $PG(2, p^n)$ are not strong except for $p^n = 2$, but that all of them have equitable main simple solutions. We shall now round out this discussion by disposing of the games $PG(2s, p^n)$ with $s \geqslant 2$.

THEOREM 7. *The games $PG(2s, p^n)$, $s \geqslant 2$, are not strong.*

Proof. Consider any $(s+1)$-space P_{s+1} in $PG(2s, p^n)$. Since any s-space intersects P_{s+1} in a space of dimension at least one, a set B will be a blocking coalition if it consists of points of P_{s+1} such that B meets every line of P_{s+1} but contains no s-space of P_{s+1}. We show that such a set B exists. Introduce a homogeneous co-ordinate system $(x_0, x_1, \ldots, x_{s+1})$ into P_{s+1} in the usual way by means of an $(s+1)$-simplex of co-ordinates σ_{s+1}.

Case 1. Suppose either $p^n \neq 2$, or $p^n = 2$ and s is even. Let B be the set of all points x of P_{s+1} such that the number $Z(x)$ of zero co-ordinates of point x satisfies $1 \leqslant Z(x) \leqslant s$.

We prove first that every line l of P_{s+1} intersects B. Clearly, l meets the s-space $x_0 = 0$ in at least one point x. If $Z(x) \neq s + 1$, it is the desired point of B. If $Z(x) = s + 1$, then let the remaining non-zero co-ordinate of x be x_i, $i \neq 0$. Let y be a point of intersection of l with the s-space $x_i = 0$. If $Z(y) \neq s + 1$, it is the desired point of B. If $Z(y) = s + 1$, then the point $x + y$ is a point of l having $Z(x + y) = s$, and is therefore in B.

We must still prove that B contains no s-space of P_{s+1}. Let an equation of an arbitrary s-space of P_{s+1} be

$$a_0x_0 + a_1x_1 + \ldots + a_{s+1}x_{s+1} = 0. \tag{8}$$

If at least one coefficient, say a_0, is equal to zero, then the point $(1, 0, 0, \ldots, 0)$ is a point of the s-space not in B. If all coefficients of (8) are different from zero, we consider two cases, $p^n = 2$ or $p^n \neq 2$. If $p^n = 2$, and s is even, then $s + 2$ is even, and, since all $a_i = 1$ by hypothesis, we have

$$\sum_{i=0}^{s+1} a_i = 0 \bmod 2;$$

hence $(1, 1, \ldots, 1)$ is a point of the s-space not in B. If $p^n \neq 2$, let $c \neq 0, 1$ and consider the numbers

$$a_0 + a_1 + \ldots + a_s \tag{9}$$

and

$$a_0 + a_1 + \ldots + a_{s-1} + c\, a_s. \tag{10}$$

At least one of these is not zero, because if both were zero then subtraction would yield $(c - 1)a_s = 0$ and hence $a_s = 0$ contrary to the hypothesis that all $a_i \neq 0$. If (9) is not zero, then the point

$$\left(1, 1, \ldots, 1, 1, -\frac{a_0 + \ldots + a_s}{a_{s+1}}\right)$$

satisfies (8) but is not in B. If (10) is not zero, then the point

$$\left(1, 1, \ldots, 1, c, -\frac{a_0 + \ldots + a_{s-1} + c\, a_s}{a_{s+1}}\right)$$

satisfies (8) but is not in B.

Case 2. Suppose $p^n = 2$, and s is odd, $s > 3$. Let B be the set of all points x of P_{s+1} with $Z(x) \neq 1, s + 1$.

We prove first that every line l of P_{s+1} intersects B. Let x be a point common to l and the s-space $x_0 = 0$. If $x \neq (0, 1, 1, \ldots, 1)$, $(0, 1, 0, \ldots, 0)$, $(0, 0, 1, 0, \ldots, 0), \ldots, (0, 0, \ldots, 0, 1)$, then $x \in B$. If x is one of these points, let y be a point common to l and the s-space $y_i = 0$, where the ith co-ordinate of x is 1, so that $x \neq y$. Suppose, for example, $i = 1$. If $y \neq (1, 0, 1, \ldots, 1)$, $(1, 0, 0, \ldots, 0)$, $(0, 0, 1, 0, \ldots, 0), \ldots, (0, 0, \ldots, 0, 1)$, then $y \in B$. It is easily seen that in any of the remaining cases, $Z(x + y) = 2, s$, or 0. Hence $x + y$ is a point of l belonging to B.

It is still necessary to prove that B contains no s-space of P_{s+1}. Let (8) be again an equation of an arbitrary s-space of P_{s+1}. If at least one coefficient, say a_0, is zero, then the point $(1, 0, \ldots, 0)$ satisfies (8) but is not in B. If all $a_i \neq 0$, then all $a_i = 1$, and, since s is odd, the point $(0, 1, 1, \ldots, 1)$ satisfies (8) but is not in B. This completes the proof.

Geometrically, in Case 1, B consists of all the points of the face-planes of dimensions $s, \ldots, 1$ of the co-ordinate simplex σ_{s+1} excluding the vertices. In Case 2, B consists of all the points of P_{s+1} excepting the vertices of σ_{s+1} and the points of the s-face-planes not lying on face-planes of lower dimension. It is not difficult to count the number of points in B and to see that this number is greater than the number of points in an s-space. But in the next theorem we shall show that this must be true for any blocking coalition in $PG(2s, p^n)$, $s \geq 2$; and hence, by Lemma 2, that there exists an equitable main simple solution.

In $PG(2s, p^n)$, let α_i be the number of points in an i-space, and let $\alpha_j{}^i$ be the number of i-spaces containing a given j-space.

Lemma 4. *If $r < s$, then $\alpha_{r-1}{}^r > 1 + \alpha_s$.*

Proof. By an easy calculation, we get

$$\alpha_{r-1}^{r} = \frac{p^{rn} + \ldots + p^{2sn}}{p^{rn}} = 1 + p^n + p^{2n} + \ldots + p^{(2s-r)n} = \alpha_{2s-r}.$$

But $r < s$ implies $2s - r > s$, or $2s - r \geq s + 1$. Hence $\alpha_{2s-r} - \alpha_s \geq \alpha_{s+1} - \alpha_s = p^{(s+1)n}$. Since $p^n > 1$, we have $\alpha_{2s-r} - \alpha_s > 1$.

Theorem 8. *The games $PG(2s, p^n)$, $s \geq 2$, have equitable main simple solutions.*

Proof. By Lemma 2 it suffices to show that if B is any blocking coalition, then the number $|B|$ of points in B is greater than α_s. Suppose, contrarywise, that $|B| \leq \alpha_s$.

If every line joining two points of B were contained in B, then B would be a t-space. If $t \geq s$, B could not be a blocking coalition since it would contain an s-space or minimal winning coalition. If $t < s$, then there would be an s-space in $PG(2s, p^n)$ not meeting B, contrary to the assumption that B is a blocking coalition. Therefore, there exists a line l with at least two points in B and at least one point x not in B.

We now prove inductively that for each $r \leq s - 1$ there exists an r-space containing x but not intersecting B. For $r = 0$, the point x suffices. Suppose the assertion is correct for $r < s - 1$. By Lemma 4, there are more than $1 + \alpha_s$ $(r + 1)$-spaces containing the given r-space; since only one of them can contain l, there are more than α_s $(r + 1)$-spaces containing the given r-space but not containing l. One of these $(r + 1)$-spaces does not intersect B, since at most one of them can meet a given point of B; for, if two of them contained the same point of B, then this point would be in the intersection of these two

$(r + 1)$-spaces which is the given r-space, contradicting the induction hypothesis that this r-space does not intersect B. This completes the induction.

In particular, there exists an $(s - 1)$-space containing x but not intersecting B. Every s-space containing this $(s - 1)$-space meets B in at least one point since B is a blocking coalition. But the s-space determined by l and the given $(s - 1)$-space meets B in at least two points. Therefore $|B| \geqslant 1 + \alpha_s$, contrary to the supposition that $|B| \leqslant \alpha_s$. This completes the proof.

7. Affine resolvable games. In this section, we examine certain simple games formed from block designs but not using all the blocks of the design as minimal winning coalitions.

A balanced design is termed *affine resolvable* if the b blocks can be divided into r classes of n blocks each, such that:

(a) every one of the classes of n blocks contains a complete replication of the v elements;

(b) any two blocks of different classes have the same number of elements in common.

Then (cf. **1**) we have $b = nr$, $v = nk$, $b = v + r - 1$, and $s_{ij} = |B_i \cap B_j| = k^2/v$ if B_i and B_j are in different classes. If we arbitrarily select one block from each class as a minimal winning coalition, we obtain a simple game, with $|\mathfrak{W}^m| = r = b/n$, which we term an *affine resolvable game*. Not all these games formable from a given affine resolvable balanced design need have the same number of players. For example, an affine resolvable balanced design with $v = 12$, $b = 22$, $r = 11$, $k = 6$, $\lambda = 5$, $n = 2$ is given (cf. **1**) by the blocks

$$
\begin{array}{ll}
B_1 = (1, 3, 4, 5, 9, 11) & B_{12} = (2, 6, 7, 8, 10, 12) \\
B_2 = (2, 4, 5, 6, 10, 1) & B_{13} = (3, 7, 8, 9, 11, 12) \\
B_3 = (3, 5, 6, 7, 11, 2) & B_{14} = (4, 8, 9, 10, 1, 12) \\
B_4 = (4, 6, 7, 8, 1, 3) & B_{15} = (5, 9, 10, 11, 2, 12) \\
B_5 = (5, 7, 8, 9, 2, 4) & B_{16} = (6, 10, 11, 1, 3, 12) \\
B_6 = (6, 8, 9, 10, 3, 5) & B_{17} = (7, 11, 1, 2, 4, 12) \\
B_7 = (7, 9, 10, 11, 4, 6) & B_{18} = (8, 1, 2, 3, 5, 12) \\
B_8 = (8, 10, 11, 1, 5, 7) & B_{19} = (9, 2, 3, 4, 6, 12) \\
B_9 = (9, 11, 1, 2, 6, 8) & B_{20} = (10, 3, 4, 5, 7, 12) \\
B_{10} = (10, 1, 2, 3, 7, 9) & B_{21} = (11, 4, 5, 6, 8, 12) \\
B_{11} = (11, 2, 3, 4, 8, 10) & B_{22} = (1, 5, 6, 7, 9, 12)
\end{array}
$$

where B_i and B_{i+11} $(i = 1, \ldots, 11)$ constitute the ith class. One affine resolvable game with eleven players has $B_i (i = 1, \ldots, 11)$ as minimal winning coalitions. Another affine resolvable game with twelve players has $B_j (j = 12, \ldots, 22)$ as minimal winning coalitions. In both cases $|B_i \cap B_j| = k^2/v = 3$ if $i \neq j$. In the first case $\{1, 3, 4\}$ is a blocking coalition; in the second case $\{12\}$ is a blocking coalition. Many other affine resolvable games can be formed from the same design.

Another example of an affine resolvable balanced design is an $EG(2, p^n)$, where the classes of blocks are the parallel pencils of lines. Selecting one line from each parallel pencil, we have an affine resolvable game with $|B_i \cap B_j| = 1$ if $i \neq j$.

As an immediate consequence of Theorem 2 and Lemma 1, we have the following theorem.

THEOREM 9. *If an affine resolvable balanced design has $k^2/v > 1$, then any affine resolvable game obtained from it as above is not strong and has no main simple solution.*

8. Interpretation in terms of linear graphs and network flows. Any simple game G can be represented as an even (or bipartite, or simple) graph, as follows. Let the two vertex sets be $\mathfrak{W}^m = \{W_1, W_2, \ldots, W_b\}$ and $N = \{1, 2, \ldots, v\}$ and let $W_i \in \mathfrak{W}^m$ and $j \in N$ be joined by an arc if and only if j is a member of W_i. Each vertex W_i has degree $|W_i|$, the number of members of W_i. The many-valued mapping $\Gamma\colon N \to \mathfrak{W}^m$, where $\Gamma(j)$ is the set of all minimal winning coalitions to which j belongs, is such that $\Gamma^{-1}W_i \cap \Gamma^{-1}W_j \neq \phi$ for $i \neq j$, or, in other words, $\Gamma\Gamma^{-1}W_i = \mathfrak{W}^m$ for each W_i. If G is a block design game then the degree of every vertex W_i of $\mathfrak{W}^m$ is k and the degree of every vertex j of N is r. To a blocking coalition of any simple game G in this representation corresponds a subset B of N such that $\Gamma B = \mathfrak{W}^m$ but $B \not\supset \Gamma^{-1} W_i$ for any $W_i \in \mathfrak{W}^m$.

We can convert this graph theoretic representation into a network flow representation as follows. Join all vertices of N to an input vertex I, and all vertices of $\mathfrak{W}^m$ to an output vertex U, as in Fig. 1, illustrating the dual of $EG(2, 2)$ with $v = 6$, $b = 4$, $k = 3$, $r = 2$ (cf. Example 1, § 10). Putting capacities c_{ij} on the arcs as indicated in the figure, a blocking coalition corresponds to a flow x_{ij} yielding maximum output but with the restriction that

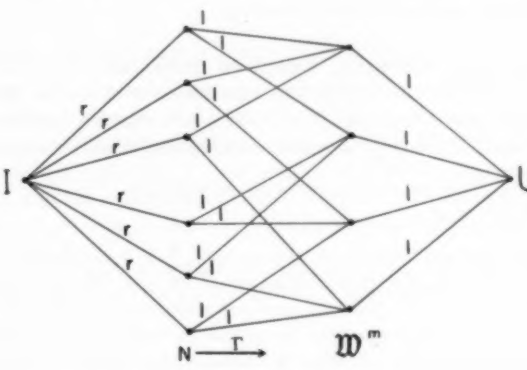

FIG. 1.

the flow shall not be different from 0 at any entire set of vertices of the form $\Gamma^{-1}W_t$, $W_t \in \mathfrak{W}^m$. This can, in turn, be expressed as a linear programming problem: find x_{ij} such that

$$\sum_{\alpha} x_{\alpha\beta} - \sum_{\gamma} x_{\beta\gamma} = 0$$

for each vertex $\beta \neq I, U$, and such that

$$\sum_{j \in \mathfrak{W}^m} x_{jU} = \max = b \tag{1}$$

subject to the constraints

$$0 \leqslant x_{ij} \leqslant c_{ij}, \tag{2}$$

with the additional restriction that

(3) for each $j \in \mathfrak{W}^m$ there exists an $i = i(j) \in \Gamma^{-1}(j)$ such that

$$x_{Ii} = \sum_j x_{ij} = 0.$$

If such a flow exists, a blocking coalition is given by the set

$$\left\{ i \in N \,\middle|\, \sum_j x_{ij} > 0 \right\}.$$

If such a flow does not exist, the game is strong. By the methods of Goldman and Tucker **(6)**, all extreme feasible vectors of the linear programme given by (1) and (2) can be determined, and then each $W_j \in \mathfrak{W}^m (j = 1, 2, \ldots, b)$ can be examined to see if the additional restriction (3) is satisfied. For if any feasible vector is on a co-ordinate $(n - r)$-plane

$$x_{i_1} = x_{i_2} = \ldots = x_{i_r} = 0$$

then so is some extreme feasible vector.

Thus the results of the preceding sections are readily interpreted in terms of linear graphs or network flows, as desired.

9. Miscellaneous remarks. Unsolved problems. In Luce **(9)**, it is proved that a necessary and sufficient condition for a simple game to be h-unstable is that there exist an $(h + 1)$-element winning coalition and that the intersection of all $(h + 1)$-element winning coalitions be empty.

THEOREM 10. *A block design game is h-stable for $1 \leqslant h < k - 1$ and h-unstable for $k - 1 \leqslant h < v - 1$.*

Proof. There exists a winning coalition with $h + 1$ members if $h \geqslant k - 1$. Clearly, for $h = k - 1$, the intersection of all $(h + 1)$-element winning coalitions is empty since $r < b$. As long as there remain two different elements to adjoin to the $(h + 1)$-element sets to obtain $(h + 2)$-element sets, induction shows that the intersection of all $(h + 1)$-element winning coalitions is empty for $h < v - 1$. This completes the proof.

We define an automorphism (collineation or perhaps cowineation) of a block design with incidence matrix $A = ||a_{ij}||$ to be a permutation π_R of the rows of A (elements or players) which carries columns of A (blocks) into columns of A. Let π_C be the permutation of the columns induced by the permutation π_R. *We shall assume that A has neither duplicated columns nor duplicated rows.* Let G_R be the group of all automorphisms π_R and let G_C be the group of all π_C.

LEMMA 5. *If π_R induces π_C then in the dual design the automorphism π_C induces π_R.*

Proof. Let π_R carry the matrix A into the matrix B. Then

$$b_{ij} = a_{\pi_R(i),j} = a_{i,\pi_C(j)}$$

for all i, j. Hence π_R is induced by π_C.

LEMMA 6. *No two different row permutations $\pi_R \neq \pi_R'$ can induce the same column permutation.*

Proof. If so, then

$$a_{i,\pi_C(j)} = a_{i,\pi'_C(j)}$$

for all i, j and hence

$$a_{\pi_R(i),j} = a_{\pi'_R(i),j}$$

for all i, j contrary to hypothesis.

THEOREM 11. *The automorphism groups of dual designs are isomorphic* (*as groups, even though of different degrees*).

Proof. Obviously, the product of two row permutations induces the product of the induced column permutations. Hence, with our restriction of non-duplication of rows and columns of A, the homomorphism is one-to-one, and onto.

The following unsolved problems seem to be difficult:

1. How can one determine the automorphism group of a block design (without examining one by one each of the permutations of the symmetric group on v letters to see if it is an automorphism, although this might be feasible within limits with a computer)? This is solved for the desarguesian finite projective spaces and the associated euclidean spaces. Further, what can be said of the transitivity of this group acting on elements and on blocks?

2. What is the minimum number of members in a blocking coalition of a block design game? This is unsolved even for finite projective planes, except for $PG(2, 3)$.

3. Do the block design games, not covered by **(13)** or the corollaries of Theorem 2, above, possess main simple solutions?

4. To determine all block design games given the parameters v, b, k, r such that $bk = vr$. That is, to determine all $v \times b$ matrices A with $a_{ij} = 0$ or 1 such that the row sums are all equal to r, the column sums are all equal to k, and the elements s_{ij} of $S = A^T A$ are all positive.

5. If $v = b$ and (hence) $k = r$, when do there exist permutation matrices P, Q such that $PAQ = A^T$ where A is the incidence matrix? When do there exist permutation matrices P, Q such that PAQ is a symmetric matrix? When is there a row permutation such that RA is symmetric? What are general criteria for permutation equivalence of matrices with elements equal to 0 or 1? for general matrices?

10. Examples. We collect in this section some concrete examples illustrating some of the preceding theorems. Other examples of block designs yielding simple games can be found in **(2)**.

Example 1. The finite euclidean plane $EG(2, 2)$ has $v^* = 4$, $b^* = 6$, $k^* = 2$, $r^* = 3$, $\lambda^* = 1$ and (Fig. 2) incidence matrix

	a	b	c	d	e	f
1	1	1	1	0	0	0
2	1	0	0	1	1	0
3	0	1	0	1	0	1
4	0	0	1	0	1	1

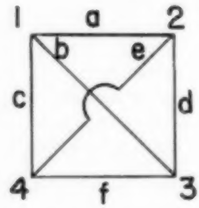

FIG. 2.

In the dual game we have $v = 6$, $b = 4$, $k = 3$, $r = 2$ and the intersection of every pair of blocks has one element. The incidence matrix of the game is the transpose of the above. The sets $\{a, f\}$, $\{b, e\}$, $\{c, d\}$ are blocking coalitions illustrating Theorem 1(a). By Lemma 3, there is no equitable main simple solution. But in this example, it is easy to give a direct proof that no main simple solution exists at all. For the linear system (6), (7) becomes, with obvious changes in notation:

$$\begin{array}{llllllll}
x_a & + x_b & + x_c & & & & = 1 \\
x_a & & & + x_d & + x_e & & = 1 \\
& x_b & & + x_d & & + x_f & = 1 \\
& & x_c & & + x_e & + x_f & = 1 \\
x_a & & & & & + x_f & > 1 \\
& x_b & & & + x_e & & > 1 \\
& & x_c & + x_d & & & > 1.
\end{array}$$

This implies $x_a = x_f > \frac{1}{2}$, $x_b = x_e > \frac{1}{2}$, $x_c = x_d > \frac{1}{2}$ so that a contradiction would result.

Example 2. The symmetric partially balanced block design designated as $R1$ in **(2)** has $v^* = b^* = 6$, $k^* = r^* = 3$, $\lambda_1 = 2$, $\lambda_2 = 1$, $n_1 = 1$, $n_2 = 4$ and incidence matrix

	a	b	c	d	e	f
1	1	0	0	1	0	1
2	1	1	0	0	1	0
3	0	1	1	0	0	1
4	1	0	1	1	0	0
5	0	1	0	1	1	0
6	0	0	1	0	1	1

The dual has $v = b = 6$, $k = r = 3$, every pair of blocks has an intersection of one or two elements, and the incidence matrix is the transpose of the above. The set $\{a, b, c\}$ is a blocking coalition. This illustrates Theorem 1.

Example 3. The symmetric regular group divisible partially balanced block design designated as $R2$ in **(2)** has $v^* = b^* = 6$, $r^* = k^* = 4$, $\lambda_1 = 3$, $\lambda_2 = 2$, $n_1 = 2$, $n_2 = 3$ and incidence matrix

	a	b	c	d	e	f
1	1	0	0	1	1	1
2	1	1	1	1	0	0
3	0	1	0	1	1	1
4	1	1	1	0	1	0
5	0	0	1	1	1	1
6	1	1	1	0	0	1

Here $s_{ij} = 2$ or 3 and indeed $\{1, 2\}$ is a blocking coalition contained in the minimal winning coalition a in accordance with Corollary (d) of Theorem 2. The dual has $v = b = 6$, $k = r = 4$, every pair of blocks has an intersection of 2 or 3 elements, and its incidence matrix is the transpose of the above. In accordance with Theorem 2, there is a blocking coalition contained (properly) in a minimal winning coalition, namely, $\{a, b, c\}$, which is a proper subset of blocks 2 or 4, or $\{a, d\}$ which is a proper subset of blocks 1 or 2. In fact the dual is isomorphic to the original design as can be seen by performing the permutations

$$\begin{pmatrix}123456\\152634\end{pmatrix} \quad \text{and} \quad \begin{pmatrix}\text{abcdef}\\\text{acedfb}\end{pmatrix}$$

on the rows and columns respectively.

Example 4. Let $EG(2, 3)$ be the euclidean plane over the integers modulo 3, with $v^* = 9$, $b^* = 12$, $k^* = 3$, $r^* = 4$, $\lambda^* = 1$. The projective plane $PG(2, 3)$ has the cyclic representation

0	1	2	3	4	5	6	7	8	9	10	11	12
1	2	3	4	5	6	7	8	9	10	11	12	0
3	4	5	6	7	8	9	10	11	12	0	1	2
9	10	11	12	0	1	2	3	4	5	6	7	8

Taking the line {12, 0, 2, 8} as line at infinity, and deleting its points, we derive the incidence matrix of $EG(2, 3)$:

	a	b	c	d	e	f	g	h	i	j	k	l
1	1	1	0	0	0	1	0	0	0	0	0	1
3	1	0	1	1	0	0	0	1	0	0	0	0
4	0	1	0	1	1	0	0	0	1	0	0	0
5	0	0	1	0	1	1	0	0	0	1	0	0
6	0	0	0	1	0	1	1	0	0	0	1	0
7	0	0	0	0	1	0	1	1	0	0	0	1
9	1	0	0	0	0	0	1	0	1	1	0	0
10	0	1	0	0	0	0	0	1	0	1	1	0
11	0	0	1	0	0	0	0	0	1	0	1	1

In the dual, we have $v = 12$, $b = 9$, $k = 4$, $r = 3$, every pair of blocks intersects in one element, and the incidence matrix is the transpose of the above. Note that $EG(2, 3)$ is also the simplest non-trivial Steiner triple system. The process of Theorem 5 applied to $\{a, b, f, l\}$, choosing $f = [1, 5, 6]$, $l = [1, 7, 11]$, yields $e = [4, 5, 7]$ and $k = [6, 10, 11]$ as substitutes, producing the blocking coalition $\{a, b, e, k\}$. But this is not minimal since $\{a, e, k\}$ is also a blocking coalition.

Example 5. One of the Steiner triple systems with $v^* = 13$ (cf. **8**) has the incidence matrix

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1	1	1	0	0
4	0	1	0	0	0	0	1	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
5	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1	1
6	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	1	0
7	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	1	1	1	0	0	0	0	0	0	0	0
8	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0	0	1	0	1	1	0	0	0	0	0	0
9	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	1	0	0	1	0	1
10	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	1
11	0	0	0	0	1	0	0	0	0	0	1	0	0	0	1	1	0	0	1	0	1	0	0	0	0	0
12	0	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	1	0	1	0	1	0	0	0	0
13	0	0	0	0	0	1	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	1	1	0

The dual has $v = 26$, $b = 13$, $k = 6$, $r = 3$, every pair of blocks intersects in one element, the incidence matrix is the transpose of the above, and $\{a, m, s, w, f\}$ is a blocking coalition.

Example 6. The design $T9$ of **(2)** has $v^* = 10$, $b^* = 6$, $k^* = 5$, $r^* = 3$, $\lambda_1 = 1$, $\lambda_2 = 2$, $n_1 = 6$, $n_2 = 3$, and incidence matrix

	a	b	c	d	e	f
1	1	1	0	0	1	0
2	0	0	1	0	1	1
3	1	0	0	1	0	1
4	0	1	1	1	0	0
5	0	1	0	1	0	1
6	0	0	1	1	1	0
7	1	0	1	0	0	1
8	1	1	1	0	0	0
9	1	0	0	1	1	0
10	0	1	0	0	1	1

The dual has $v = 6$, $b = 10$, $k = 3$, $r = 5$, and the incidence matrix is the transpose of the above. This game is strong, and has a main simple vector $a_i = 1/3$ $(i = 1, 2, \ldots, 6)$.

Example 7. The design SR14 of **(2)** is a symmetric semi-regular group divisible design with $v = b = 9$, $k = r = 6$, $\lambda_1 = 3$, $\lambda_2 = 4$, and incidence matrix:

	a	b	c	d	e	f	g	h	i
1	1	1	1	1	1	1	0	0	0
2	1	0	1	1	0	1	1	0	1
3	1	0	1	0	1	1	1	1	0
4	1	1	1	0	0	0	1	1	1
5	1	1	0	1	1	0	1	1	0
6	1	1	0	1	0	1	0	1	1
7	0	0	0	1	1	1	1	1	1
8	0	1	1	0	1	1	0	1	1
9	0	1	1	1	1	0	1	0	1

Here all $s_{ij} \geqslant \lambda_1 = 3$, and indeed $\{1, 2, 3\}$ is a blocking coalition contained in the minimal winning coalition a, as promised by Corollary (f) of Theorem 2.

References

1. R. C. Bose, *A note on the resolvability of balanced incomplete block designs*, Sankhyā, *6* (1942), 105–110.
2. R. C. Bose, W. H. Clatworthy, S. S. Shrikhande, *Tables of partially balanced designs with two associate classes*, Tech. Bull. No. 107, North Carolina Agricultural Experiment Station (1954).
3. R. C. Bose and K. R. Nair, *Partially balanced incomplete block designs*, Sankhyā, *4* (1939), 337–372.
4. R. D. Carmichael, *Introduction to the theory of groups of finite order* (Ginn, 1937).
5. W. S. Connor, *Some relations among the blocks of symmetrical group divisible designs*, Ann. Math. Stat., *23* (1952), 602–609.
6. A. J. Goldman, A. W. Tucker, *Theory of linear programming. In* H. W. Kuhn and A. W. Tucker, *Linear inequalities and related systems* (Princeton, 1956).
7. H. M. Gurk and J. R. Isbell, *Simple solutions. In* A. W. Tucker and R. D. Luce, *Contributions to the theory of games IV* (Princeton, 1959).
8. M. Hall, Jr., *A survey of combinatorial analysis. In* I. Kaplansky, E. Hewitt, M. Hall, Jr., and R. Fortet, *Some aspects of analysis and probability* (Wiley, 1958).
9. R. D. Luce, *A definition of k-stability for n-person games*, Ann. Math., *59* (1954), 357–366.
10. E. H. Moore, *Tactical memoranda*, Amer. J. Math., *18* (1896), 264–303.
11. E. Netto, *Lehrbuch der Combinatorik* (Chelsea reprint of 1927 edition).
12. J. von Neumann and O. Morgenstern, *Theory of games and economic behavior* (2nd ed.; Princeton, 1947).
13. M. Richardson, *On finite projective games*, Proc. Amer. Math. Soc., *7* (1956), 458–465.
14. L. S. Shapley, *Lectures on n-person games* (Princeton University notes, unpublished).
15. S. S. Shrikhande, *On the dual of some balanced incomplete block designs*, Biometrica, *8* (1952), 66–72.

General Electric Company
and
Brooklyn College and Princeton University

SIMPLE ALGEBRAS OF TYPE (1, 1) ARE ASSOCIATIVE

ERWIN KLEINFELD

1. Introduction. In the classification of almost alternative algebras relative quasiequivalence by Albert two new classes of algebras of type (γ, δ) were introduced, namely those of type $(1, 1)$ and $(-1, 0)$ (**1**, equations (34), (35), and Theorem 6). Since rings of type $(1, 1)$ and $(-1, 0)$ are anti-isomorphic it suffices to consider those of type $(1, 1)$. They may be defined as rings satisfying

$$B(x, y, z) = (x, y, z) - (x, z, y) = 0, \tag{1}$$

and

$$A(x, y, z) = (x, y, z) + (y, z, x) + (z, x, y) = 0, \tag{2}$$

for all elements x, y, and z of the ring, where the associator (a, b, c) is given by $(a, b, c) = (ab)c - a(bc)$.

Actually the identity

$$(x, x, x) = 0, \tag{2'}$$

together with (1) implies (2) whenever the characteristic of the ring is different from 2. This may readily be verified by linearizing (2′). Consequently we may use (1) and (2′) as the defining relations for a ring of type $(1, 1)$.

Additional results on rings of type $(1, 1)$ were obtained by Kokoris (**3**; **4**) and the author (**2**). The main result of the present paper, which is stated in the title, draws upon these results. Let R be a ring of type $(1, 1)$, u any element of R of the form $u = (x, y, x)$, and C the right ideal of R generated by u. Then $uC = 0 = Cu$ (Theorem 2). This turns out to be the key result in the structure theory for it assures the existence of an abundance of right ideals even under the assumption of simplicity (Theorems 6 and 8). In contrast to this, if R is also not associative then it has no proper left ideals (Theorem 4). Every minimal right ideal A of R has the property $A^2 = 0$ (Theorem 5). With the additional hypothesis of chain conditions on right ideals no maximal right ideal of R can be nil and the union of the minimal right ideals of R is contained in the intersection of the maximal right ideals (Theorem 8). By assuming either that R has an idempotent or that R is a finite dimensional algebra one can utilize the information about the right ideals of R in order to reach a contradiction. In fact even primitive rings and hence semi-simple rings of type $(1, 1)$ turn out to be associative [Theorem 11 and its Corollaries]. The characteristic of R is assumed to be different from 2 and 3, and in §4 different from 5 as well.

Received June 9, 1959. The research for this paper was supported in part by a grant from the Office of Ordnance Research to Ohio State University.

We also consider the more general question of rings of type (γ, δ). When $\gamma \neq 1, -1$ it turns out that a simple ring that is not associative, has no proper left or right ideals, and therefore the techniques developed for rings of type $(1, 1)$ are not applicable.

2. Identities. Fundamental to all our results on rings of type $(1, 1)$ is Theorem 2, already mentioned in the Introduction, which permits the construction of right ideals. This result must be obtained through complicated computation. It is a more sophisticated version of the identity $(x, y, x)^2 = 0$, which constituted the main result of **(2)**. We shall save considerable time and effort by recalling the following identities that hold for all elements w, x, y, z of a ring of type $(1, 1)$. Except for (10), which is a specialization of (1) and (2), these identities may easily be located in **(2)**. The commutator (x, y) is defined by $(x, y) = xy - yx$.

$$(3) \qquad (x, (x, y, z)) = 0,$$

$$(4) \qquad C(w, x, y, z) = (w, (x, y, z)) + (x, (w, y, z)) = 0,$$

$$(5) \qquad D(x, y, z) = (x, yz) - y(x, z) - (x, y)z + (x, y, z) = 0,$$

$$(6) \qquad F(w, x, y, z) = (wx, y, z) - (w, xy, z) + (w, x, yz) - w(x, y, z) - (w, x, y)z = 0,$$

$$(7) \qquad H(w, x, y, z) = (w, (w, x, y)z) - (w, (w, x, z)y) = 0,$$

$$(8) \qquad ((x, y, z), x, x) = 0,$$

$$(9) \qquad ((x, y, x), y, x) = ((x, y, x), y) = ((x, y, x), yx) = 0,$$

$$(10) \qquad -(y, x, x) = 2(x, y, x) = 2(x, x, y).$$

In addition to these identities we shall also make use of a result of Kokoris **(4)**, that every subring of a ring of type $(1, 1)$ that is generated by a single element is associative.

Our first objective is to establish the following generalization of (3).

THEOREM 1. *Let v be any element of the form $v = (w, x, y)$, where w, x, y are elements of a ring R of type* $(1, 1)$. *Then the right ideal D generated by v has the property that* $(w, D) = 0$.

Proof. Let $s = (y, x, x)$ and $u = (x, y, x)$. Then $-s = 2u$, as a result of (10). Since $(u, yx) = 0$ is implied by (9), we know that $(s, yx) = 0$. Also

$$0 = C(y, yx, x, x) = (y, (yx, x, x)) + (yx, s) = (y, (yx, x, x)).$$

But

$$\begin{aligned} 0 &= F(y, x, x, x) \\ &= (yx, x, x) - (y, x^2, x) + (y, x, x^2) - y(x, x, x) - (y, x, x)x \\ &= (yx, x, x) - (y, x, x)x, \end{aligned}$$

as a result of (1). Since $(y,(yx, x, x)) = 0$, we must also have $(y,(y, x, x)x) = 0$. Replacing x by $x + z$ in this last identity and utilizing (1) and (7), it must follow that $3(y,(y, x, x)z) + 3(y,(y, z, z)x) = 0$. Now replacing x by $-x$ in this last identity and adding one obtains $6(y,(y, x, x)z) = 0$. Because of our assumption on the characteristic of R we may divide by 6, so that $(y,(y, x, x)z) = 0$. At this point replace x by $x + w$. Then utilization of (1) results in $(y,(y, w, x)z) = 0$. In summary, we have shown that

$$(w, (w, x, y)p) = 0. \tag{11}$$

From here on we consider elements of the form

$$(w, x, y)R_pR_qR_{z_3} \ldots R_{z_n}$$

where R_k is the mapping $a \rightarrow ak$. Our inductive assumption will be that w commutes with all such elements for a given n and we shall attempt to prove this for $n + 1$. Incidentally (11) suffices to start off the induction. In case $n = 2$ we merely leave off

$$T = R_{z_3}R_{z_4} \ldots R_{z_{n+1}}.$$

Consider therefore $t = (w, [(w, x, y)p \cdot q]T)$. In attempting to show that $t = 0$, the first step consists of establishing that the value of t is unchanged under all permutations on x, y, p, q. That the interchange of x and y does not alter the value of t is a consequence of (1). Starting with

$$0 = F(w, x, y, p) = (wx, y, p) - (w, xy, p) + (w, x, yp) - w(x, y, p) - (w, x, y)p,$$

we apply the mapping R_qT to this equation and commute the result with w. Because of the induction hypothesis we have

$$(w, (w, xy, p)R_qT) = 0 = (w, (w, x, yp)R_qT).$$

Therefore

$$(w, (wx, y, p)R_qT) - (w, [w(x, y, p)]R_qT) - (w, [(w, x, y)p \cdot q]T) = 0.$$

In the first two terms of this last identity one may interchange y and p without changing their values, hence this must be true of the third term. But that term is $-t$. Finally the induction hypothesis implies $(w, [(w, x, y) \cdot pq]T) = 0$, so that

$$t = (w, [(w, x, y)p \cdot q]T) = (w, ((w, x, y), p, q)T).$$

But in the last term p and q may be permuted without change in value, so that this must also be true of t. This suffices to demonstrate that every permutation of x, y, p, q leaves t unchanged. Suppose now that

$$t' = (w, [(w, x, x)x \cdot x]T).$$

Because of (8) and (10), $t' = (w, [(w, x, x) \cdot x^2]T)$, and the latter is zero as a result of the induction hypothesis. Therefore $t' = 0$. In t' replace x by $x + p$

and also by $x - p$ and add the two expressions. Now, utilizing the fact that every permutation of x, y, p, q in t does not alter the value of t, we see that $12(w, [(w, x, x)p \cdot p]T) = 0$, so that $(w, [(w, x, x)p \cdot p]T) = 0$. By replacing x by $x + y$ in the last identity and then replacing p by $p + q$ it becomes clear that $(w, [(w, x, y)p \cdot q]T) = 0$. This of course completes the induction. However, (w, D) consists of sums of elements that are of the same type as t, but where n is arbitrary. Consequently $(w, D) = 0$. This completes the proof of the theorem.

At this point we are ready to prove the basic

THEOREM 2. *Let u be any element of the form $u = (x, y, x)$, where x and y are elements of a ring R of type* $(1, 1)$. *Then the right ideal C generated by u has the property that $uC = 0 = Cu$.*

Proof. Let c be an arbitrary element of C. Then Theorem 1 implies that $(c, x) = 0$. Because of (10), $(y, x, x) = -2(x, y, x) = -2u$, so that the right ideal generated by (y, x, x) must also be C. A second application of Theorem 1 yields that $(c, y) = 0$. Suppose that

$$T = R_{z_1} R_{z_2} \dots R_{z_n}.$$

Then as a consequence of Theorem 1 it follows that $((r, a, b)T, r) = 0$. If we replace r by $r + s$ in this last identity it becomes clear that

$$((s, a, b)T, r) = -((r, a, b)T, s). \tag{12}$$

Suppose then that we set $r = yx$, $s = y$, and $a = b = x$ in (12). Then we obtain $((y, x, x)T, yx) = -((yx, x, x)T, y)$. However, it follows from

$$\begin{aligned} 0 &= F(y, x, x, x) \\ &= (yx, x, x) - (y, x^2, x) + (y, x, x^2) - y(x, x, x) - (y, x, x)x \\ &= (yx, x, x) - (y, x, x)x, \end{aligned}$$

that $[(y, x, x)x]T = (yx, x, x)T$. Since $[(y, x, x)x]T$ is an element of C and $(C, y) = 0$, it must be that $((xy, x, x)T, y) = 0$. But then it follows from above that $((y, x, x)T, yx) = 0$. In other words $(c, yx) = 0$, because every element of C may be written as a sum of elements of the form $(y, x, x)T$. We have already established that $(c, y) = (c, x) = 0$, so that

$$0 = D(c, y, x) = (c, yx) - y(c, x) - (c, y)x + (c, y, x) = (c, y, x).$$

But then $(c, x, y) = 0$, as a result of (1). Expanding

$$0 = D(c, x, y) = (c, xy) - x(c, y) - (c, x)y + (c, x, y) = (c, xy),$$

it also follows that $(c, xy) = 0$. In summary, we have shown that

$$(c, xy) = (c, yx) = (c, x) = (c, y) = (c, x, y) = (c, y, x) = 0. \tag{13}$$

Also setting $r = u$, $s = y$, and $a = b = x$ in (12) we find that $((y, x, x)T, u) =$

$-((u, x, x)T, y)$. But $(u, x, x) = 0$, as a result of (8), so that $((y, x, x)T, u) = 0$. From this one may conclude as before that

$$(c, u) = 0. \tag{14}$$

Since $u = (xy)x - x(yx)$ it follows from (14) that $-(c, xy\cdot x) + (c, x\cdot yx) = 0$. But then

$$\begin{aligned} 0 &= -D(c, xy, x) + D(c, x, yx) \\ &= -(c, xy\cdot x) + xy\cdot(c, x) + (c, xy)x - (c, xy, x) \\ &\qquad + (c, x\cdot yx) - x(c, yx) - (c, x)\cdot yx + (c, x, yx) \\ &= (c, x, yx) - (c, xy, x), \end{aligned}$$

as a result of (13). But then the last identity may be used in

$$0 = F(c, x, y, x) = (cx, y, x) - (c, xy, x) + (c, x, yx) - c(x, y, x) - (c, x, y)x,$$

together with (13) and the observation that cx is an element of C, to establish $-c(x, y, x) = 0$. This implies that $cu = 0$. But then as a result of (14) we also must have $uc = 0$. This argument holds for every element c of C, so that $uC = 0 = Cu$. This completes the proof of the theorem.

3. The structure of left and right ideals. In this section R will be assumed to be a simple ring of type (1, 1), of characteristic different from 2 and 3, that is, not associative. In this connection simple means that the only two-sided ideals of R are either R or 0. This hypothesis on R may of course lead to a contradiction, in which case we would be justified in concluding that simple rings of type (1, 1) are associative. Indeed we obtain this result in §4 with the added assumption that R is a finite dimensional algebra. In the present section we shall adhere to the more general situation.

THEOREM 3. *Rings of type* (1, 1) *that have no proper right ideals are associative.*

Proof. Form $u = (x, y, x)$, for arbitrary elements x and y of R. Let C be the right ideal generated by u. Then either $C = 0$, in which case $u = 0$, or $C = R$. In the latter case we may make use of Theorem 2 in order to obtain that $uR = 0 = Ru$. The set of all elements q of R with the property that $qR = 0 = Rq$ may be verified to form a two-sided ideal of R. Since R is simple, either all such q are zero, or $R^2 = 0$. In the latter instance R would be associative, contrary to assumption. Therefore in all cases $u = (x, y, x) = 0$. But then (1) and (2) may be employed to establish that $(y, x, x) = 0$. Replacing x by $x + z$ in this last identity we are forced to concludeth at R is associative, a contradiction. The contradiction is the result of the assumption that R was not associative. This concludes the proof of the theorem.

The situation on left ideals is just the reverse.

THEOREM 4. *Simple rings of type* (1, 1) *that are not associative have no proper left ideals.*

Proof. Let B be a proper left ideal of R. An element s of B will be defined to be special (relative to B) in case sR is always contained in B. It is easy to verify that the set S of special elements is closed under subtraction. It turns out we can even show that the special elements form a two-sided ideal of R. Select arbitrary elements x, y, z in R, a, b in B, and s in S. Then $(x, y, b) = (xy)b - x(yb)$ is an element of the left ideal B. Then because of (1), (x, b, y) must also be in B. But then (b, x, y) is also in B, as a result of (2). Since s is in B and B is a left ideal of R it follows that xs is in B. On the other hand $(xs)y = (x, s, y) + x(sy)$. But it follows from the definition of S that sy must be in B, so that $x(sy)$ is also in B. Since s is in B it follows from previous discussion that (x, s, y) is in B. As a result both xs and $(xs)y$ are in B, hence xs is special. Then we know that S is a left ideal of R. In a similar manner sx is in B because of the definition of S, while $(sx)y = (s, x, y) + s(xy)$ is also in B. Then S is a two-sided ideal of R. Since S is contained in B and B is a proper left ideal of R, we must have $S \neq R$. But R is simple, so that $S = 0$. This is very useful information since we can show that various elements of R must be special. Finally we shall be able to deduce that $B = 0$, which is the desired contradiction. To begin with (a, b, x) is clearly in B. Also

$$0 = F(a, b, x, y) = (ab, x, y) - (a, bx, y) + (a, b, xy) - a(b, x, x) - (a, b, x)y,$$

so that $-(a, b, x)y$ is also in B. Hence (a, b, x) is special and consequently $(a, b, x) = 0$. In other words $(B, B, R) = 0$. Then from (1) and (2) it follows that also $(B, R, B) = 0$ and $(R, B, B) = 0$. For this reason

$$\begin{aligned} 0 &= F(x, y, b, b) \\ &= (xy, b, b) - (x, yb, b) + (x, y, b^2) - x(y, b, b) - (x, y, b)b \\ &= (x, y, b^2) - (x, y, b)b, \end{aligned}$$

so that $(x, y, b^2) = (x, y, b)b$. On the other hand

$$\begin{aligned} 0 &= F(x, b, b, y) + F(x, b, y, b) \\ &= (xb, b, y) - (x, b^2, y) + (x, b, by) - x(b, b, y) - (x, b, b)y \\ &\qquad + (xb, y, b) - (x, by, b) + (x, b, yb) - x(b, y, b) - (x, b, y)b \\ &= -(x, b^2, y) + (x, b, by) - (x, by, b) - (x, b, y)b. \end{aligned}$$

But $(x, b, by) - (x, by, b) = 0$ because of (1), while $(x, b^2, y) = (x, y, b^2)$, and $(x, b, y)b = (x, y, b)b$, for the same reason. But then $(x, y, b^2) = -(x, y, b)b$, whereas we have already noted that $(x, y, b^2) = (x, y, b)b$. Therefore $(x, y, b^2) = (x, y, b)b = 0$. The nucleus N of R is defined as the set of all elements n in R such that $(n, R, R) = (R, n, R) = (R, R, n) = 0$. Because of (1) and (2), b^2 must be in N. Let n be an arbitrary element of N. Then

$$0 = C(n, x, y, z) = (n, (x, y, z)) + (x, (n, y, z)) = (n, (x, y, z)),$$

so that $(n, (R, R, R)) = 0$. Similarly (12) implies that $(n, (R, R, R)R) = 0$. However, the set of finite sums of elements that are of the form (R, R, R)

and of the form $(R, R, R)R$ form a two-sided ideal in an arbitrary ring R. If that ring is also simple and not associative then of course it follows that this ideal must be the whole ring. The conclusion we can draw in the present situation is that $(n, R) = 0$, so that $(b^2, R) = 0$. Since b^2 is an element of B and now also $b^2x = xb^2$ is an element of B we conclude that b^2 must be special, hence $b^2 = 0$. If we replace b by $a + b$ then also $ab + ba = 0$. Since (x, y, b) is an element of B it must then follow that $(x, y, b)a = -a(x, y, b)$. Also

$$\begin{aligned} 0 &= C(a, x, y, b) \\ &= (a, (x, y, b)) + (x, (a, y, b)) = (a, (x, y, b)) \\ &= a(x, y, b) - (x, y, b)a. \end{aligned}$$

But then $a(x, y, b) = (x, y, b)a = 0$. This may be used in the expansion of

$$0 = F(x, y, b, a) = (xy, b, a) - (x, yb, a) + (x, y, ba) - x(y, b, a) - (x, y, b)a,$$

to show that $(x, y, ba) = 0$. As before this implies ba is in the nucleus, and hence commutes with every element of R. Consequently one can show that ba is special and so $ba = 0$. In other words $B^2 = 0$. Form $I = B + BR$. Then $(bx)y = (b, x, y) + b(xy)$. We have already noted that (b, x, y) is an element of B, and hence of I. Then I must be a right ideal of R. Similarly $y(bx) = -(y, b, x) + (yb)x$, where $-(y, b, x)$ is an element of B and $(yb)x$ is an element of BR. Thus $y(bx)$ is an element of I. This suffices to show that I is a two-sided ideal of R. If $I = 0$, then $B = 0$, contrary to assumption. If on the other hand $I = R$, then $BI = B(B + BR) = 0$, so that $BR = 0$. But then $I = B$, and so $R = B$, contrary to assumption. In either case we have reached a contradiction. Consequently there can exist no proper left ideal B in R. This completes the proof of the theorem.

If Theorem 4 were true for right ideals also, then of course this would prove simple rings of type (1, 1) to be associative, which is the strongest possible result one could hope to get. Arguments of the type used in the proof of Theorem 4 seem inadequate for this purpose. With some effort one can obtain a construction for a right ideal of R, properly contained within any non-zero right ideal A of R. Except when A is minimal, this construction does not seem to be especially enlightening, and since we can get some information on minimal right ideals more directly, we shall not go into this construction.

THEOREM 5. *If R is a simple ring of type* (1, 1) *that is not associative, and if A is a minimal right ideal of R, then $A^2 = 0$.*

Proof. Let $t = (a, x, x)$, where a is an arbitrary element of A and x an arbitrary element of R. Let C denote the right ideal generated by t. Since t is an element of A, C must also be contained in A. From the minimality of A as a right ideal it follows that either $C = 0$, or that $C = A$. If $C = 0$ then $t = 0$. On the other hand Theorem 2 implies that $Ct = 0 = tC$, so that if

$C = A$, then $At = 0 = tA$. In either case we may conclude that $At = 0 = tA$. Replacing x by $x + y$ in this last identity and using (1) we conclude that $A(a, x, y) = 0 = (a, x, y)A$, where y is an arbitrary element of R. In other words $A(A, R, R) = 0 = (A, R, R)A$. Let P be the set of all elements p in A with the property that $Ap = A(pR) = 0$. P is obviously closed under subtraction. With the last identity we shall be able to show that P is in fact a right ideal of R. Select arbitrary elements p in P, a, b, d in A, and x, y, z in R. Since p is in A, so is px. From the definition of P it follows that $a(px) = 0$. Also $a(px \cdot y) = a(p, x, y) + a(p \cdot xy) = a(p, x, y)$. Since $0 = A(A, R, R)$, we have $a(p, x, y) = 0$, and thereby $a(px \cdot y) = 0$. Consequently P is a right ideal of R and contained in A. Let us assume that $A^2 \neq 0$. Then $P \neq A$. From the minimality of A as a right ideal it follows that $P = 0$. Our next objective will be to show that (A, A, R) is contained in P. We have seen previously that (A, A, R) is contained in A and also that $A(A, R, R) = 0$. But then

$$
\begin{aligned}
0 &= aF(b, d, x, y) \\
&= a(bd, x, y) - a(b, dx, y) + a(b, d, xy) - a[b(d, x, y)] - a[(b, d, x)y] \\
&= -a[(b, d, x)y],
\end{aligned}
$$

and so (A, A, R) is contained in P. But then $(A, A, R) = 0$. As a consequence of the last identity

$$
\begin{aligned}
0 = F(a, b, x, y) = (ab, x, y) - (a, bx, y) + (a, b, xy) - a(b, x, y) \\
- (a, b, x)y = (ab, x, y).
\end{aligned}
$$

In other words $(A^2, R, R) = 0$. As in the proof of Theorem 4 we can now use $0 = C(ab, x, y, z)$ to show that $(A^2, (R, R, R)) = 0$, and (12) to show that $(A^2, (R, R, R)R) = 0$. Since the two-sided ideal generated by all associators must be all of R, we have demonstrated that $(A^2, R) = 0$. But then $x(ab) = (ab)x = (a, b, x) + a(bx) = a(bx)$, proving that A^2 is a two-sided ideal of R. Since we have assumed that $A^2 \neq 0$, it must be the case that $A^2 = R$. But A^2 is contained in A, so that $A = R$. This contradicts the assumption that A is a minimal right ideal of R. Consequently $A^2 = 0$. This completes the proof of the theorem.

The next result plays a very important part in the Main Theorem.

THEOREM 6. *Let R be a simple ring of type* (1, 1), *with chain conditions on right ideals, that is not associative. Then the number of maximal right ideals and the number of minimal right ideals are both greater than one.*

Proof. The existence of at least one maximal right ideal and of at least one minimal right ideal are insured by the chain conditions and Theorem 3. Suppose that R has only one maximal right ideal. Call it B. Consider an arbitrary element u of the form $u = (y, x, x)$, and let C be the right ideal generated by u. Then, because of Theorem 2, $uC = 0 = Cu$. If u is not an

element of B, then C is not contained in B. But B is the unique maximal right ideal of R and there is only one right ideal not contained in it, namely R. Thence $uR = 0 = Ru$. But the absolute divisors of zero of R form a two-sided ideal of R, which cannot be all of R. Consequently $u = 0$, contrary to assumption. Thus u is an element of B for all x and y in R. Replace x by $x + z$ in u.

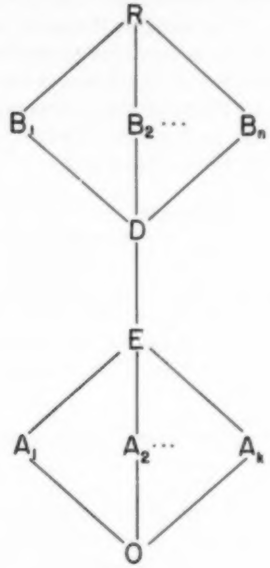

FIG. 1

Then as a result of (1) we note that all associators of R must be contained in B and thereby also all right multiples of associators. But then $B = R$, a contradiction since B was assumed to be a maximal right ideal. Because of this contradiction B cannot be the unique maximal right ideal and consequently R must have at least two maximal right ideals. This completes the first half of the theorem. Suppose now that A is the only minimal right ideal of R. Define u as before, as well as C, so that $uC = 0 = Cu$. We see at once that either C contains A or C must be zero. In the latter case $u = 0$. In the former $uA = 0 = Au$. But then we may conclude that $uA = 0 = Au$ in either case. Replacing x by $x + z$ in the last identity and using (1) we obtain $(R, R, R)A = 0 = A(R, R, R)$. Let w, x, y, z be arbitrary elements of R and a an arbitrary element of A. Then

$$\begin{aligned} 0 &= F(w, x, y, z)a \\ &= (wx, y, z)a - (w, xy, z)a + (w, x, yz)a - [w(x, y, z)]a - [(w, x, y)z]a \\ &= -[w(x, y, z)]a - [(w, x, y)z]a. \end{aligned}$$

Therefore in $q = [w(x, y, z)]a$ all permutations of x, y, and z do not alter the value of q. But $0 = (x, y, z) + (y, z, x) + (z, x, y)$, because of (2). Therefore $3q = 0$, so that $q = 0$. But then $[(w, x, y)z]a = 0$. In summary, we have shown that $(R, R, R)A = 0$, and that $[(R, R, R)R]A = 0$. As before we can deduce from this $RA = 0$, so that $A = 0$. However, A was chosen to be a minimal right ideal and this is clearly a contradiction. Hence R has at least two minimal right ideals. This concludes the proof of the theorem.

The last result seems to leave open the possibility that a maximal right ideal might be a minimal right ideal. However we shall see later that this cannot happen. In fact every minimal right ideal will be seen to be contained in every maximal right ideal, and any such pair are always separated by at least one intermediate right ideal (Theorem 8).

THEOREM 7. *Let R be a simple ring of type* (1, 1). *Suppose A and B are right ideals of R such that $A^2 = 0$, $A + B = R$, and $B \neq R$. Then R is associative.*

Proof. Since $A^2 = 0$, we see that $(A, A, R) = 0$. But then $(A, R, A) = 0$ and $(R, A, A) = 0$ as a result of (1) and (2). Also (B, R, R) is contained in B and therefore so is (R, B, B), as a result of (2). Expanding

$$\begin{aligned}(R, R, R) &= (A + B, A + B, A + B)\\ &= (A, A, A) + (B, B, B) + (A, B, B) + (B, A, B) + (B, B, A)\\ &\qquad + (B, A, A) + (A, B, A) + (A, A, B),\end{aligned}$$

it becomes evident that (R, R, R) is contained in B. But then also $(R, R, R)R$ is contained in B. Since $B \neq R$, the only two-sided ideal of R that is contained in B is zero. But we have just seen that the ideal generated by all associators is contained in B. Therefore R must be associative. This completes the proof of the theorem.

COROLLARY. *Let R be a simple ring of type* (1, 1) *that is not associative. If A is a minimal right ideal of R and B a maximal right ideal of R then A is contained in B.*

Proof. Suppose that A is not contained in B. Since B is a maximal right ideal of R, $A + B = R$. Since A is a minimal right ideal of R, we may use Theorem 5 to obtain $A^2 = 0$. But then the hypothesis of Theorem 7 is satisfied, so that R must be associative. From this contradiction one deduces that A is contained in B. This completes the proof of the corollary.

THEOREM 8. *Let R be a simple ring of type* (1, 1) *with unit element and chain conditions on right ideals that is not associative. Let B be any maximal right ideal of R, A any minimal right ideal of R, D the intersection of all the maximal right ideals of R, and E the union of all the minimal right ideals of R. Then B is is not nil and*

$$0 \subset A \subset E \subseteqq D \subset B \subset R.$$

Proof. Suppose that B is nil (that means every element of B is nilpotent). Theorem 6 implies the existence of another maximal right ideal $B' \neq B$. Therefore $B + B' = R$. Since the unit element 1 is in R, there must exist elements x in B and y in B', such that $1 = x + y$. Then $1 - x = y$. Suppose that $x^n = 0$. Let $s = 1 + x + \ldots + x^{n-1}$. Then $(1 - x)s = 1 = ys$. But this implies that 1 is in B', so that $B' = R$, contrary to assumption. Thus B cannot be nil. On the other hand Theorem 5 tells us that $A^2 = 0$, so that A is certainly nil. Therefore $B \neq A$. Because of Theorem 6, $E \neq A$ and $D \neq B$. Clearly A is contained in E, and D is contained in B. From the Corollary to Theorem 7 it follows that A is contained in B and hence in D. But then E must be contained in B. So far all the inclusions have been proper. However the best we can say about E and D is that E is contained in D, but in this case we are unable to eliminate the possibility that $E = D$. This completes the proof of the theorem.

Fig. 1 indicates the simplest possible structure of any ring R satisfying the hypothesis of Theorem 8, if indeed such a ring exists. The B_i denote maximal right ideals and the A_j minimal right ideals. D and E are defined in the statement of Theorem 8.

4. Main section. We shall make use of the following theorem, whose proof appears in **(4)**.

THEOREM 9 (Kokoris). *Let R be a simple ring of type* (1, 1) *that is not associative, and e any idempotent of R. Then e must be the unit element* 1 *of R.*

There appears to be a minor gap in Kokoris' proof, but fortunately a simple permutation of the facts already proved in **(4)** can be used to prove Theorem 9. We proceed with the details. In the proof of his Lemma 3, the element $a = xy$, where x is in R_{10} and y is in R_{11} is not the most general element of G_0. Rather G_0 consists of sums of such elements and hence one can only say that G_0 is the sum of nilpotent elements rather than that G_0 is nil. Let us consider the case when R is simple and $H = R$. Then $R_{11} = R_{01}R_{00}$. Moreover, it is proved that R_{11} commutes with $R_{01}R_{00}$, so that R_{11} is commutative. Now the fact that R_{11} is the sum of nilpotent elements suffices to establish that R_{11} is nil, and this of course contradicts the fact that e is in R_{11}.

We shall also make use of the following result about algebras of type (1, 1) (understood to be finite dimensional), whose proof may be found in **(1)**.

THEOREM 10 (Albert). *A nil algebra of type* (1, 1) *is nilpotent. With this background we are ready to prove the result stated in the title of the present paper.*

MAIN THEOREM. *Simple algebras of type* (1, 1) *are associative.*

Proof. Let R be a simple algebra of type (1, 1) that is not associative. We shall attempt to show that R satisfies the hypothesis of Theorem 8, but not one of the conclusions, thus obtaining the necessary contradiction. If R were nil then it would be nilpotent. Since R^2 is an ideal of R, either $R^2 = R$ or $R^2 =$

0. If $R^2 = 0$, then R would be associative, contrary to assumption. On the other hand $R^2 = R$ would contradict the fact that R is nilpotent. Therefore R is not nil. Suppose x is some element of R that is not nilpotent. The subalgebra S of R that is generated by x therefore cannot be nil. Since S is a finite dimensional, associative algebra it must contain an idempotent e. But then Theorem 9 implies that $e = 1$. Thus R contains a unit element. Since R is a finite dimensional algebra with unit element it clearly has ascending and descending chain conditions on right ideals. Thus R satisfies the hypothesis of Theorem 8. Let B be any maximal right ideal of R. If 1 were an element in B then we would have $B = R$, a contradiction. Hence 1 is not an element of B. Let y be an arbitrary element of B and T the subalgebra generated by y. If T were not nil then it would have to contain an idempotent. However, that is impossible since Theorem 9 limits any idempotent in B to be 1, and we have already seen that 1 is not in B. Thus T is nilpotent, which implies that B is nil. We have reached a contradiction since one of the conclusions of Theorem 8 states that B cannot be nil. The contradiction arose from the assumption that R was not associative. Therefore R is associative. This concludes the proof of the theorem.

Once it is known that simple algebras of type (1, 1) are associative, it is easy to extend this result to semi-simple algebras. The radical may be defined as the maximal nil ideal. One such argument follows closely the one given in **(3)** for algebras of type (γ, δ), where $\gamma \neq 1, -1$, and need not be repeated here.

At this point we shall demonstrate how the main theorem carries over to a large extent to rings without finiteness assumptions. This also results in a second and somewhat more direct proof of the main theorem (Corollary 3 of Theorem 11). As usual a ring R is defined to be primitive in case it has a maximal right ideal A, which contains no two-sided ideal of R other than the zero ideal and in case there exists an element e in R such that $ex - x$ is always in A for all x in R.

THEOREM 11. *If R is a primitive ring of type* (1, 1) *then R is associative.*

COROLLARY 1. *If R is a semi-simple ring of type* (1, 1) *then R is associative.*

COROLLARY 2. *If R is a simple ring of type* (1, 1) *and contains an idempotent then R is associative.*

COROLLARY 3. *If R is a simple, finite dimensional algebra of type* (1, 1) *then R is associative.*

Proof. Let A be a regular maximal right ideal of R which contains no two-sided ideal of R other than the zero ideal and assume that R is not associative. We assert that there exists at least one element u of the form $u = (x, y, x)$ which is not contained in A. For assume otherwise. Then (y, x, x) must also be in A. Replacing x by $x + z$ it then follows that $2(y, z, x)$ is in A, for all

elements x, y and z in R. Now it is well known, and can easily be verified directly, that in an arbitrary ring all finite sums of elements of the form (R, R, R) and $(R, R, R)R$ form a two-sided ideal I of R. In this instance I would be contained in A. But then by assumption we would have $I = 0$, and R would be associative. This is clearly a contradiction. Hence there must exist an element $u = (x, y, x)$ which is not in A. Let C be the right ideal generated by u. Since A is a maximal right ideal it follows that $A + C = R$. Then we can find an element a in A and an element c in C such that $e = a + c$. Forming $eu - u = au + cu - u$, we note that $cu = 0$ as a result of Theorem 2, while $eu - u$ is in A. Therefore $au - u$ must be an element of A. Since A is a right ideal au belongs to A, hence u must also be in A. But this is clearly a contradiction, since we deliberately chose u not in A. Hence R must have been associative to begin with. This completes the proof of the theorem.

Making use of the Jacobson-Brown radical of a ring it is clear that a semi-simple ring is a subdirect sum of primitive rings, so that Corollary 1 follows at once from the theorem.

If R is a simple, non-associative ring of type (1, 1) and contains an idempotent, then as a result of Theorem 9 R must contain a unit element 1. But then form a maximal right ideal not containing 1. This must indeed be a regular, maximal right ideal of R. It contains no ideal of R other than zero since R is simple. Therefore R is primitive, and hence associative as a result of the theorem. This is a contradiction. Hence R must have been associative to begin with. This establishes Corollary 2.

If R is a simple, finite dimensional algebra of type (1, 1) then, as in the early part of the proof of the main theorem, R is either associative or contains an idempotent. Then one may use Corollary 2 in order to establish Corollary 3.

Rings of type (1, 1) with radical need not be associative, of course. In fact it is not difficult to construct finite dimensional algebras of type (1, 1) which are not associative. It is worth noting that there exists a division ring of characteristic 2 which satisfies both (1) and (2′), yet is not associative **(5)**.

5. Rings of type (γ, δ), $\gamma \neq 1, -1$. A ring is said to be of type (γ, δ) in case identities (2), and (15), which follows, hold. Identity (15) is given by

$$J(x, y, z) = \gamma(x, z, y) + \delta(y, z, x) + (z, x, y) = 0, \tag{15}$$

where x, y, z are arbitrary elements of the ring and γ, δ are constant scalar elements. One may also assume that $\gamma^2 = \delta^2 - \delta + 1$, for otherwise one can readily verify the ring to be associative. Therefore the condition that $\gamma \neq 1$, -1 is equivalent to the condition that $\delta \neq 0.1$. In the remainder of this section we shall consider rings R of type (γ, δ), $\gamma \neq 1$, -1 whose characteristic is different from 2 and 3. We shall first develop some essential identities. As was shown in **(2)**,

$$(16)\quad G(w, x, y, z) = (w, (x, y, z)) - (x, (y, z, w)) + (y, (z, w, x)) - (z, (w, x, y)) = 0,$$

and since this may be proved in a ring satisfying (2) only, it must be satisfied by all elements of R. From

$$\begin{aligned} 0 &= G(y, x, x, x) + (x, A(x, y, x)) \\ &= (y, (x, x, x)) - (x, (x, x, y)) + (x, (x, y, x)) - (x, (y, x, x)) \\ &\qquad + (x, (x, y, x) + (x, (y, x, x)) + (x, (x, x, y)) \\ &= 2(x, y, x)), \end{aligned}$$

it follows that $(x, (x, y, x)) = 0$. But then

$$\begin{aligned} 0 &= (x, J(x, x, y)) \\ &= \gamma(x, (x, y, x)) + \delta(x, (x, y, x)) + (x, (y, x, x)) \\ &= (x, (y, x, x)). \end{aligned}$$

But then

$$\begin{aligned} 0 &= (x, A(x, y, x)) \\ &= (x, (x, y, x)) + (x, (y, x, x)) + (x, (x, x, y)) \\ &= (x, (x, x, y)). \end{aligned}$$

If in the last two identities we replace x by $x + z$ and $x - z$ and add, we obtain

$$(17)\quad K(x, y, z) = (x, (y, x, z)) + (x, (y, z, x)) + (z, (y, x, x)) = 0,$$

and

$$(18)\quad L(x, y, z) = (x, (z, x, y)) + (x, (x, z, y)) + (z, (x, x, y)) = 0.$$

Now let

$$t = (x, (x, y, z)) + (x, (x, z, y)), \text{ and } u = (x, (z, y, x)) + (x, (y, z, x)).$$

Then

$$\begin{aligned} 0 &= G(x, x, y, z) - K(x, z, y) + L(x, y, z) \\ &= (x, (x, y, z)) - (x, (y, z, x)) + (y, (z, x, x)) - (z, (x, x, y)) \\ &\qquad - (x, (z, x, y)) - (x, (z, y, x)) - (y, (z, x, x)) + (x, (z, x, y)) \\ &\qquad + (x, (x, z, y)) + (z, (x, x, y)) \\ &= (x, (x, y, z)) + (x, (x, z, y)) - (x, (z, y, x)) - (x, (y, z, x)) \\ &= t - u. \end{aligned}$$

Consequently $t = u$. On the other hand

$$\begin{aligned} 0 &= J(y, x, z) + J(z, x, y) \\ &= \gamma(y, z, x) + \delta(x, z, y) + (z, y, x) + \gamma(z, y, x) + \delta(x, y, z) + (y, z, x) \\ &= (\gamma + 1)[(z, y, x) + (y, z, x)] + \delta[(x, y, z) + (x, z, y)]. \end{aligned}$$

Commuting both sides with x one obtains $(\gamma + \delta + 1)t = 0$, since $t = u$. However

$$\begin{aligned} 0 &= J(x, y, z) + J(x, z, y) - A(x, y, z) - A(x, z, y) \\ &= \gamma(x, z, y) + \delta(y, z, x) + (z, x, y) + \gamma(x, y, z) + \delta(z, y, x) \\ &\qquad + (y, x, z) - (x, y, z) - (y, z, x) - (z, x, y) - (x, z, y) \\ &\qquad - (z, y, x) - (y, x, z) \\ &= (\gamma - 1)[(x, y, z) + (x, z, y)] + (\delta - 1)[(z, y, x) + (y, z, x)]. \end{aligned}$$

Commuting both sides with x one obtains $(\gamma + \delta - 2)t = 0$. Since both $(\gamma + \delta + 1)t = 0$, and $(\gamma + \delta - 2)t = 0$, it must be that $3t = 0$, and so $t = 0$. Since $u = t$, we also have $u = 0$. We have shown that

$$(x, (x, y, z)) + (x, (x, z, y)) = 0 = (x, (z, y, x)) + (x, (y, z, x)). \tag{19}$$

Incidentally up to this point we have made no use of the restriction on γ. However, the next result makes use of this assumption.

THEOREM 12. *Let R be a simple ring of type (γ, δ), $\gamma \neq 1, -1$ that is not associative. Then R has no proper left or right ideals.*

Proof. Let B be any proper right ideal of R. Define S as the set of all elements s of B with the property that Rs is always contained in B. Let x, y, z denote arbitrary elements of R, a, b arbitrary elements of B and s an arbitrary element of S. Since B is a right ideal of R, (b, x, y) must be an element of B. But then

$$\begin{aligned} 0 &= J(b, y, x) - A(b, y, x) \\ &= \gamma(b, x, y) + \delta(y, x, b) + (x, b, y) - (b, y, x) - (y, x, b) - (x, b, y) \\ &= (\delta - 1)(y, x, b) + \gamma(b, x, y) - (b, y, x), \end{aligned}$$

so that $(\delta - 1)(y, x, b)$ is in B. Since $\delta \neq 1$, this implies that (y, x, b) is in B. Expanding $0 = A(b, x, y)$, we note that also (y, b, x) is in B. Clearly S is closed under subtraction. We now show that in fact S is an ideal of B. Since s is in B, sy will be also. Then $z(sy) = -(z, s, y) + (zs)y$. We have already noted that (z, s, y) is in B. Also it follows from the definition of S that zs is in B. Since B is a right ideal of R, $(zs)y$ must be in B. Thereby $z(sy)$ is also in B. But this implies that sy is in S, so that S is a right ideal of R. Again from the definition of S it follows that ys is in B. Then $z(ys) = -(z, y, s) + (zy)s$, and so $z(ys)$ is also in B. But then ys is in S and therefore S is a two-sided ideal of R. However, B is a proper right ideal of R and S is contained in B. Consequently, since R is simple, $S = 0$. Next we proceed to show that a number of elements are zero by virtue of the fact that we can prove they are contained in S. Thus $0 = F(x, y, a, b) = (xy, a, b) - (x, ya, b) + (x, y, ab) - x(y, a, b) - (x, y, a)b$, implies that $-x(y, a, b)$ is contained in B. But then (y, a, b) is contained in S and hence $(R, B, B) = 0$. But then

$$\begin{aligned} 0 &= J(x, a, b) - A(x, a, b) \\ &= \gamma(x, b, a) + \delta(a, b, x) + (b, x, a) - (x, a, b) - (a, b, x) - (b, x, a) \\ &= (\delta - 1)(a, b, x). \end{aligned}$$

Since $\delta \neq 1$, $(B, B, R) = 0$. At this point

$$0 = A(x, a, b) = (x, a, b) + (a, b, x) + (b, x, a) = (b, x, a),$$

so that $(B, R, B) = 0$. In summary, we have shown that

$$(B, B, R) = (B, R, B) = (R, B, B) = 0. \tag{20}$$

Set $x = b$, $y = z = x$ in (19). Then we obtain $(b, (b, x, x)) + (b, (b, x, x)) = 0$. Hence $(b, (b, x, x)) = 0$. We shall now establish that

$$(b^2, x, x) = b(b, x, x) = (b, x, x)b. \tag{21}$$

So far we have been able to show that the second and third terms of (21) are equal.

$$\begin{aligned} 0 &= F(b, b, x, x) \\ &= (b^2, x, x) - (b, bx, x) + (b, b, x^2) - b(b, x, x) - (b, b, x)x \\ &= (b^2, x, x) - b(b, x, x), \end{aligned}$$

because of (20). Thus the first term of (21) is equal to the second term. This establishes (21). Since

$$\begin{aligned} 0 &= J(x, y, x) \\ &= \gamma(x, x, y) + \delta(y, x, x) + (x, x, y) \\ &= (\gamma + 1)(x, x, y) + \delta(y, x, x) \end{aligned}$$

and $\gamma \neq -1$, we have

$$(x, x, y) = -\left(\frac{\delta}{\gamma + 1}\right)(y, x, x).$$

But then substituting $y = b^2$ we obtain

$$(x, x, b^2) = -\left(\frac{\delta}{\gamma + 1}\right)(b^2, x, x).$$

Similarly, substituting $y = b$,

$$(x, x, b) = -\left(\frac{\delta}{\gamma + 1}\right)(b, x, x)$$

and so

$$(x, x, b)b = -\left(\frac{\delta}{\gamma + 1}\right)(b, x, x)b.$$

But we have already seen in (21) that $(b^2, x, x) = (b, x, x)b$. Then we may conclude that $(x, x, b^2) = (x, x, b)b$. Expanding

$$\begin{aligned} 0 &= F(x, x, b, b) \\ &= (x^2, b, b) - (x, xb, b) + (x, x, b^2) - x(x, b, b) - (x, x, b)b, \end{aligned}$$

we see that $-(x, xb, b) = 0$, as a result of the previous identity and (20). But then

$$\begin{aligned} 0 &= J(b, x, xb) \\ &= \gamma(b, xb, x) + \delta(x, xb, b) + (xb, b, x) \\ &= \gamma(b, xb, x) + (xb, b, x). \end{aligned}$$

However,

$$\begin{aligned} 0 &= F(b, x, b, x) \\ &= (bx, b, x) - (b, xb, x) + (b, x, bx) - b(x, b, x) - (b, x, b)x \\ &= -(b, xb, x) - b(x, b, x). \end{aligned}$$

This implies that $(b, xb, x) = -b(x, b, x)$. But then $-\gamma b(x, b, x) + (xb, b, x) = 0$. From

$$\begin{aligned} 0 &= F(x, b, b, x) \\ &= (xb, b, x) - (x, b^2, x) + (x, b, bx) - x(b, b, x) - (x, b, b)x \\ &= (xb, b, x) - (x, b^2, x) \end{aligned}$$

it follows that $(xb, b, x) = (x, b^2, x)$. Thus $-\gamma b(x, b, x) + (x, b^2, x) = 0$. In

$$\begin{aligned} 0 &= J(x, x, y) \\ &= \gamma(x, y, x) + \delta(x, y, x) + (y, x, x) \\ &= (\gamma + \delta)(x, y, x) + (y, x, x), \end{aligned}$$

substitute $y = b^2$ to obtain $(\gamma + \delta)(x, b^2, x) + (b^2, x, x) = 0$, and also $(\gamma + \delta)b(x, b, x) + b(b, x, x) = 0$. We have already established in (21) that $(b^2, x, x) = b(b, x, x)$, so that $(\gamma + \delta)(x, b^2, x) = (\gamma + \delta)b(x, b, x)$. If $\gamma + \delta = 0$, then substituting in $\gamma^2 = \delta^2 - \delta + 1$ we see that $\delta = 1$, contrary to assumption. Therefore $(x, b^2, x) = b(x, b, x)$. Since $-\gamma b(x, b, x) + (x, b^2, x) = 0$ has already been established, we combine the last two identities and get $(1 - \gamma) b(x, b, x) = 0$. But $\gamma \neq 1$, so that $b(x, b, x) = 0$, and hence all the terms in (21) are zero. Replacing x by $x + y$ in our last identity we see that

$$b(x, b, y) = -b(y, b, x). \tag{22}$$

We showed that $(x, xb, b) = 0$ earlier in the proof. On the other hand

$$\begin{aligned} 0 &= F(x, b, b, x) \\ &= (xb, b, x) - (x, b^2, x) + (x, b, bx) - x(b, b, x) - (x, b, b)x \\ &= (xb, b, x). \end{aligned}$$

But then

$$\begin{aligned} 0 &= J(xb, b, x) - \delta A(b, x, xb) \\ &= \gamma(xb, x, b) + \delta(b, x, xb) + (x, xb, b) - \delta(b, x, xb) - \delta(x, xb, b) \\ &\qquad - \delta(xb, b, x) \\ &= \gamma(xb, x, b). \end{aligned}$$

Since $\gamma \neq 0$, $(xb, x, b) = 0$. Substituting $x + y$ for x in this last identity we get

(23) $$(xb, y, b) = -(yb, x, b).$$

In the second part of (19) replace x by $w + x$, so that

$$(w, (z, y, x)) + (x, (z, y, w)) + (x, (y, z, w)) + (w, (y, z, x)) = 0.$$

Now let $w = z = b$. Then because of (20), $(b, (b, y, x)) + (b, (y, b, x)) = 0$. From (19) and (2) one proves that $(z, (y, z, x)) + (z, (x, z, y)) = 0$. Then if we let $z = b$, $(b, (y, b, x)) + (b, (x, b, y)) = 0$. But then

$$\begin{aligned} 0 &= (b, J(x, y, b)) \\ &= (b, \gamma(x, b, y) + \delta(y, b, x) + (b, x, y)) \\ &= (\gamma - \delta - 1)(b, (x, b, y)). \end{aligned}$$

If $\gamma = \delta + 1$ and $\gamma^2 = \delta^2 - \delta + 1$ then $3\delta = 0$ so that $\delta = 0$, contrary to assumption. Therefore $(b, (x, b, y)) = 0$. From this it follows readily that

(24) $$(b, (b, y, x)) = 0 = (b, (y, b, x)).$$

Then

$$\begin{aligned} 0 &= F(b, y, x, b) \\ &= (by, x, b) - (b, yx, b) + (b, y, xb) - b(y, x, b) - (b, y, x)b \\ &= (b, y, xb) - b(y, x, b) - (b, y, x)b. \end{aligned}$$

Now using (24),

$$b(y, x, b) + (b, y, x)b = b(y, x, b) + b(b, y, x) = bJ(y, x, b) - b(x, b, y).$$

Hence $(b, y, xb) = -b(x, b, y) = b(y, b, x)$, using (22). We have demonstrated that

(25) $$(b, y, xb) = b(y, b, x).$$

Now

$$\begin{aligned} 0 &= F(y, b, x, b) \\ &= (yb, x, b) - (y, bx, b) + (y, b, xb) - y(b, x, b) - (y, b, x)b \\ &= (yb, x, b) + (y, b, xb) - (y, b, x)b. \end{aligned}$$

Therefore $(yb, x, b) + (y, b, xb) = (y, b, x)b = b(y, b, x)$, as a result of (24). But then $(y, b, xb) = -(yb, x, b) + b(y, b, x) = (xb, y, b) - b(x, b, y)$, because of (23) and (22). We have demonstrated that

(26) $$(y, b, xb) = (xb, y, b) - b(x, b, y).$$

Now

$$\begin{aligned} 0 &= F(b, x, b, y) + A(xb, y, b) \\ &= (bx, b, y) - (b, xb, y) + (b, x, by) - b(x, b, y) \\ &\qquad - (b, x, b)y + (xb, y, b) + (y, b, xb) + (b, xb, y) \\ &= -b(x, b, y) + (xb, y, b) + (y, b, xb). \end{aligned}$$

But then comparison of the last identity with (26) shows that

(27) $$(xb, y, b) = b(x, b, y),$$

and

$$(y, b, xb) = 0. \tag{28}$$

From $0 = J(b, xb, y) = \gamma(b, y, xb) + \delta(xb, y, b) + (y, b, xb)$, we get $\gamma(b, y, xb) + \delta(xb, y, b) = 0$, using (28). But then as a result of (22), (25), and (27) $(\gamma - \delta)b(y, b, x) = 0$. Since $\gamma \neq \delta$, then $b(y, b, x) = 0$. As before one may also deduce $b(b, x, y) = 0$, by use of (2) and (15). But then

$$\begin{aligned} 0 &= F(b, b, x, y) \\ &= (b^2, x, y) - (b, bx, y) + (b, b, xy) - b(b, x, y) - (b, b, x)y \\ &= (b^2, x, y). \end{aligned}$$

Again using (2) and (15) one may deduce that $(x, y, b^2) = (y, b^2, x) = 0$. In other words b^2 must lie in the nucleus N of R. Furthermore it follows from an argument presented in the Appendix of **(4)** that therefore $(b^2, R) = 0$. But then we have b^2 in B and $xb^2 = b^2x$ is also in B, so that b^2 is in S. Therefore $b^2 = 0$. Replacing b by $a + b$ we see that

$$ab + ba = 0. \tag{29}$$

Now

$$\begin{aligned} 0 &= K(x, a, b) \\ &= (x, (a, x, b)) + (x, (a, b, x)) + (b, (a, x, x)) \\ &= (b, (a, x, x)). \end{aligned}$$

But then $(b, (x, a, x)) = 0$, as a result of (2) and (15). On the other hand (x, a, x) is in B, so that $b(x, a, x) = -(x, a, x)b$, using (29). Consequently

$$b(x, a, x) = 0 = (x, a, x)b. \tag{30}$$

Then

$$\begin{aligned} 0 &= F(x, a, x, b) \\ &= (xa, x, b) - (x, ax, b) + (x, a, xb) - x(a, x, b) - (x, a, x)b \\ &= (xa, x, b) + (x, a, xb). \end{aligned}$$

As a result of substituting $y = x$ in (28) we obtain $(x, b, xb) = 0$. At this point replace b by $a + b$ in the last identity. Then one gets $(x, a, xb) = -(x, b, xa)$. Therefore $(xa, x, b) = (x, b, xa)$. Adding to the last identity

$$0 = A(xa, x, b) = (xa, x, b) + (x, b, xa) + (b, xa, x),$$

we get

$$2(xa, x, b) + (b, xa, x) = 0.$$

From

$$\begin{aligned} 0 &= F(b, x, a, x) \\ &= (bx, a, x) - (b, xa, x) + (b, x, ax) - b(x, a, x) - (b, x, a)x \\ &= -(b, xa, x) \end{aligned}$$

one may now deduce that $2(xa, x, b) = 0$, so that $(xa, x, b) = 0$. But we have

noted previously that $(xa, x, b) + (x, a, xb) = 0$. Hence

$$(x, a, xb) = 0. \tag{31}$$

We note that

$$\begin{aligned} 0 &= F(x, a, y, b) \\ &= (xa, y, b) - (x, ay, b) + (x, a, yb) - x(a, y, b) - (x, a, y)b \\ &= (xa, y, b) + (x, a, yb) - (x, a, y)b. \end{aligned}$$

However, replacing x by $x + y$ in (31) we see that $(x, a, yb) = -(y, a, xb)$. Replacing b by $a + b$ in (28) shows that $-(y, a, xb) = (y, b, xa)$. Therefore $(x, a, yb) = (y, b, xa)$. Consequently $(xa, y, b) + (y, b, xa) = (x, a, y)b$. Subtracting $0 = A(xa, y, b) = (xa, y, b) + (y, b, xa) + (b, xa, y)$ from the last equation we see that $-(b, xa, y) = (x, a, y)b$. Because of (29) it follows that $(x, a, y)b = -b(x, a, y)$, and thereby $(b, xa, y) = b(x, a, y)$. Comparing the last identity with

$$\begin{aligned} 0 &= F(b, x, a, y) \\ &= (bx, a, y) - (b, xa, y) + (b, x, ay) - b(x, a, y) - (b, x, a)y \\ &= -(b, xa, y) - b(x, a, y), \end{aligned}$$

we conclude that $(b, xa, y) = b(x, a, y) = 0$. Consequently $(B, RB, R) = 0$. Define $D = B + RB$. It is a simple matter to verify that D is a two-sided ideal of R. Since D contains B, a non-trivial right ideal of R, it must be that $D = R$. Therefore $(B, RB, R) = 0$ and $(B, B, R) = 0$ imply $(B, R, R) = (B, D, R) = 0$. But then B is contained in the nucleus N of R and as before then $(B, R) = 0$ follows from **(2)**. This, however, suffices to show that B is contained in S, so that $B = 0$. But this is clearly a contradiction. It arose out of the original assumption that B was a proper right ideal of R. Therefore R can have no proper right ideals. The argument that R can have no proper left ideals follows from the fact that a ring of type (γ, δ) is anti-isomorphic to one of type $(-\gamma, 1 - \delta)$ **(4**, Theorem 1**)**. This completes the proof of the theorem.

We have purposely omitted from our discussion the rings of type $(-1, 1)$ and their anti-isomorphic copies, the rings of type $(1, 0)$. The former are right alternative rings that satisfy (2). From the structure theory of right alternative algebras it follows that simple algebras of type $(-1, 1)$ whose characteristic is different from 2 and 3 are associative. This result is analogous to our Main Theorem. Rings of type $(-1, 1)$ have been considered by Maneri for his PhD. dissertation. His results will be published elsewhere.

References

1. A. A. Albert, *Almost alternative algebras*, Portugal. Math., *8* (1949), 23–36.

2. E. Kleinfeld, *Rings of* (γ, δ) *type*, Portugal. Math., *18* (1959), 107–110.

3. L. A. Kokoris, *On a class of almost alternative algebras*, Can. J. Math., *8* (1956), 250–255.

4. ——— *On rings of* (γ, δ) *type*, Proc. Amer. Math. Soc., *9* (1958), 897–904.

5. R. L. San Soucie, *Right alternative division rings of characteristic 2*, Proc. Amer. Math. Soc., *6* (1955), 291–296.

Ohio State University

THE GROUPS OF REGULAR COMPLEX POLYGONS

D. W. CROWE

1. Introduction. The two-dimensional unitary space, $\mathbf{U}_2$, is a complex vector space of points $(x, y) = (x_1 + ix_2, y_1 + iy_2)$, for which the distance between (x, y) and (x', y') is defined by $[(x - x')\overline{(x - x')} + (y - y')\overline{(y - y')}]^{\frac{1}{2}}$. A *unitary transformation* is a linear transformation which preserves distance. A *line* is the set of points (x, y) satisfying some complex equation $ax + by = c$. A unitary transformation is a (*unitary*) *reflection* if it is of finite period $n > 1$ and leaves a line pointwise invariant. Thus a unitary matrix represents a reflection if its two characteristic roots are 1 and a complex nth root $(n > 1)$ of 1.

Shephard **(7)** has introduced the notion of *regular complex polygon* as follows. Consider a configuration **P** consisting of points ("vertices") and lines ("edges") in $\mathbf{U}_2$. If the group of automorphisms of **P** is generated by two reflections, one, say S, which permutes cyclically the vertices on an edge and another, T, which permutes cyclically the edges at one of these vertices, then **P** is called a regular complex polygon. Now the finite groups in $\mathbf{U}_2$ generated by S and T can be interpreted as finite groups of orthogonal transformations in four-dimensional Euclidean space, $\mathbf{E}_4$. These groups in $\mathbf{E}_4$ have been enumerated by Seifert and Threlfall **(6)**, using the fact that each such group is homomorphic (either 2:1 or 1:1) to one of the finite groups of displacements in elliptic space of three dimensions enumerated by Goursat **(5)**. The purpose of this paper is to find the groups in Goursat's list corresponding to Shephard's groups.

In §2 we find quaternion transformations $q' = aqb$ corresponding to the generators of Shephard's groups. In §3 these are used to associate groups $\mathfrak{L}$ and $\mathfrak{R}$ of Clifford translations to Shephard's groups. (This discussion closely follows that of **(6)**.) In §4 Goursat's groups are described analogously, leading to the natural homomorphism between Shephard's groups and Goursat's described in §5. The results are tabulated, and summarized in the Theorem.

We write $\mathfrak{C}_n$ and $\mathfrak{E}$ for cyclic groups of order n and 1 respectively. The polyhedral group defined by $A^\mu = B^\nu = (AB)^2 = E$ is denoted $(2, \mu, \nu)$, and the binary polyhedral group $A^\mu = B^\nu = (AB)^2$ is $\langle 2, \mu, \nu \rangle$. With $\mathfrak{C}_n$ the latter are the only finite groups of quaternions. For quaternions the exponential form $\exp s\pi\mathbf{j}/n$ means $\cos s\pi/n + \mathbf{j} \sin s\pi/n$. The order of a group $\mathfrak{G}$ is $|\mathfrak{G}|$.

2. The quaternion representation of a unitary reflection. If the

Received June 29, 1959. This paper is a portion of the author's Ph.D. thesis at the University of Michigan, prepared under the direction of Professor H. S. M. Coxeter.

point $(x, y) = (x_1 + ix_2, y_1 + iy_2)$ in $\mathbf{U}_2$ is represented by the point (x_1, x_2, y_1, y_2) in $\mathbf{E}_4$, then the transformation

$$(2.1) \qquad (x_1' + ix_2', y_1' + iy_2') = (x, y) \begin{pmatrix} r_1 + ir_2 & s_1 + is_2 \\ t_1 + it_2 & u_1 + iu_2 \end{pmatrix}$$

is represented by the transformation

$$(2.2) \qquad (x_1', x_2', y_1', y_2',) = (x_1, x_2, y_1, y_2) \begin{pmatrix} r_1 & r_2 & s_1 & s_2 \\ -r_2 & r_1 & -s_2 & s_1 \\ t_1 & t_2 & u_1 & u_2 \\ -t_2 & t_1 & -u_2 & u_1 \end{pmatrix}.$$

In particular, if 2.1 is a unitary reflection then 2.2 is proper orthogonal. The transformation 2.2 can in turn be expressed as a quaternion transformation **(2)**

$$(2.3) \qquad q' = (a_1 + \mathbf{i}a_2 + \mathbf{j}a_3 + \mathbf{k}a_4)\, q\, (b_1 + \mathbf{i}b_2 + \mathbf{j}b_3 + \mathbf{k}b_4),$$

where $q = x_1 + \mathbf{i}x_2 + \mathbf{j}y_1 + \mathbf{k}y_2$, $q' = x_1' + \mathbf{i}x_2' + \mathbf{j}y_1' + \mathbf{k}y_2'$, and $Na = Nb = 1$. Since in our case 2.2 corresponds to a unitary reflection, we have also $a_1 = b_1$ **(2**, p. 141**)**.

The a_i and b_i in 2.3 can be found in terms of the r_i, s_i, t_i, and u_i in 2.2 by applying 2.2 and 2.3 to $(1, 0, 0, 0)$, $(0, 1, 0, 0)$, $(0, 0, 1, 0)$, $(0, 0, 0, 1)$ and $\mathbf{1}$, $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ respectively, and equating coefficients. For example, applying 2.2 and 2.3 to $(1, 0, 0, 0)$ and $\mathbf{1}$ yields

$$r_1 = a_1b_1 - a_2b_2 - a_3b_3 - a_4b_4,\ r_2 = a_1b_2 + a_2b_1 + a_3b_4 - a_4b_3,$$
$$s_1 = a_1b_3 + a_3b_1 - a_2b_4 + a_4b_2,\ s_2 = a_1b_4 + a_4b_1 + a_2b_3 - a_3b_2.$$

Repeating this in the other three cases yields 12 more equations. Adding and subtracting these 16 equations in pairs containing $r_1, r_2, \ldots, u_1, u_2$ yields 16 equivalent equations which reduce to

$$\begin{gathered} a_3 = a_4 = 0 \text{ and} \\ 2a_1b_1 = r_1 + u_1,\ 2a_1b_3 = s_1 - t_1, \\ 2a_1b_2 = r_2 - u_2,\ 2a_1b_4 = s_2 + t_2, \\ 2a_2b_1 = r_2 + u_2,\ 2a_2b_3 = s_2 - t_2, \\ 2a_2b_2 = u_1 - r_1,\ 2a_2b_4 = -s_1 - t_1. \end{gathered}$$

These equations readily give the quaternion transformation 2.3 corresponding to a unitary matrix. For example, if

$$T = \begin{pmatrix} 1 & 0 \\ 0 & \exp 2\pi i/n \end{pmatrix}$$

we get

$$a_1b_3 = a_2b_3 = a_1b_4 = a_2b_4 = 0.$$

Since either a_1 or a_2 is different from zero this implies $b_3 = b_4 = 0$. Moreover, $2a_1b_2 = -u_2 = -2a_2b_1$, that is

$$a_1/a_2 = -b_1/b_2,$$

and since

$$a_1b_1 - a_2b_2 = r_1 > 0$$

we have $b = \bar{a}$. Then

$$2a_1b_1 = 2a_1^2 = r_1 + u_1 = 1 + \cos 2\pi/n,$$

so that

$$a_1 = \pm \cos \pi/n = b_1.$$

We choose the upper sign. Finally

$$2a_2b_1 = 2a_2a_1 = 2a_2 \cos \pi/n = \sin 2\pi/n,$$

which yields

$$a_2 = \sin \pi/n = -b_2.$$

Consequently the quaternion form of T is

$$q' = (\cos \pi/n + \mathbf{i} \sin \pi/n)\, q\, (\cos \pi/n - \mathbf{i} \sin \pi/n). \tag{2.4}$$

In Table I at the end of this paper we list the groups of the regular complex polygons, writing $p_1[t]p_2$ for the group of the polygon $p_1\{t\}p_2$ in the notation of (**4**, p. 80). The generators S and T are taken from (**7**) except for the group $2[n]2$ for which the given S is not a reflection. For this group we let

$$S = S^{-1} = \begin{pmatrix} \cos 2\pi/n & \sin 2\pi/n \\ \sin 2\pi/n & -\cos 2\pi/n \end{pmatrix} \quad \text{and} \quad T = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Professor Coxeter has pointed out that the defining relations for Shephard's groups are particularly simple in terms of S^{-1} and T. Consequently we use these generators in preference to S and T. The quaternion form of S^{-1} appears in the table. By 2.4, T is always $(\exp \pi\mathbf{i}/p_2)\, q\, (\exp -\pi\mathbf{i}/p_2)$.

3. The groups $\mathfrak{L}$, $\mathfrak{R}$, $\mathfrak{l}$, $\mathfrak{r}$. If S^{-1} is represented by $q' = aqb$ and T by $q' = cqd$ we designate by $\mathfrak{L}^*$ the group generated by a and c, and by $\mathfrak{R}^*$ the group generated by b and d. These groups are either cyclic groups or binary polyhedral groups $\langle 2, \mu, \nu\rangle$ (**1**), and can thus be readily determined. We remark that in all cases $\mathfrak{L}^*$ is cyclic.

We turn now to a more detailed discussion of Shephard's groups in terms of the groups $\mathfrak{L}^*$ and $\mathfrak{R}^*$. Let $\mathfrak{G}$ be such a group. Then $\mathfrak{G}$ is a group of transformations $q' = aqb$, $a \in \mathfrak{L}^*$, $b \in \mathfrak{R}^*$. But not every transformation of the given form is in the group, and furthermore there are certain redundancies. The latter occur because the transformation $q' = aqb$ is identical to the transformation $q' = (-a)q(-b)$. There are no other redundancies of this nature, for if $aqb = cqd$ for all q then $c^{-1}a = db^{-1}$ is a real number, say s, so that $c = as^{-1}$ and $d = sb$. But $Na = Nb = Nc = Nd = 1$, so $s = \pm 1$. We remove these redundancies by identifying the elements (a, b) and $(-a, -b)$ of $\mathfrak{L}^* \times \mathfrak{R}^*$. (Observe that multiplication in $\mathfrak{L}^* \times \mathfrak{R}^*$ is defined by (a, b) $(c, d) = (ac, bd)$, which does indeed correspond to composition of the corres-

ponding transformations since the commutativity in $\mathfrak{L}^*$ implies $ac = ca$.) For finite groups of quaternions the only case in which this identification is trivial is if either $\mathfrak{L}^*$ or $\mathfrak{R}^*$ is a cyclic group of odd order. For our groups this is never the case. Essentially we thus form $(\mathfrak{L}^* \times \mathfrak{R}^*)/\mathfrak{C}_2$, whose elements are the classes $\{(a, b)\}$. We let

$$\mathfrak{L} = \{\{(a,\ 1)\} : a \in \mathfrak{L}^*\} \text{ and } \mathfrak{R} = \{\{(1, b)\} : b \in \mathfrak{R}^*\}.$$

The elements of $\mathfrak{L}$ and $\mathfrak{R}$ can be multiplied in the obvious manner:

$$\{(a, 1)\}\ \{(1, b)\} = \{(a, b)\}.$$

Clearly $\mathfrak{R} \cong \mathfrak{R}^*$ and $\mathfrak{L} \cong \mathfrak{L}^*$, but $\mathfrak{L}\mathfrak{R} \cong (\mathfrak{L}^* \times \mathfrak{R}^*)/\mathfrak{C}_2$. Every element of $\mathfrak{L}$ commutes with every element of $\mathfrak{R}$. The group $\mathfrak{G}$ is isomorphic to a subgroup of $\mathfrak{L}\mathfrak{R}$ and will be treated as if it were itself a subgroup.

Let $\mathfrak{l} = \mathfrak{L} \cap \mathfrak{G}$ and $\mathfrak{r} = \mathfrak{R} \cap \mathfrak{G}$. Then $\mathfrak{l}$ is a normal subgroup of $\mathfrak{L}$. For let $L \in \mathfrak{L}$ and $l \in \mathfrak{l}$. Certainly $L^{-1}lL \in \mathfrak{L}$. There is some $R \in \mathfrak{R}$ such that $LR \in \mathfrak{G}$, for $\mathfrak{L}$ consists of exactly such elements L. Consequently

$$L^{-1}l\,L = (LR)^{-1}l\,(LR) \in \mathfrak{G}.$$

That is, $L^{-1}\,lL \in \mathfrak{L} \cap \mathfrak{G} = \mathfrak{l}$. Similarly, $\mathfrak{r}$ is a normal subgroup of $\mathfrak{R}$. Furthermore $\mathfrak{l}\mathfrak{r}$ is a normal subgroup in $\mathfrak{G}$ of order $\frac{1}{2}|\mathfrak{l}||\mathfrak{r}|$. For if $l \in \mathfrak{l}$ and $r \in \mathfrak{r}$ then $(LR)^{-1}lr\,(LR) = (L^{-1}l\,L)\ (R^{-1}r\,R) \in \mathfrak{l}\mathfrak{r}$.

We say that an element $L \in \mathfrak{L}$ is *paired* with an element $R \in \mathfrak{R}$ if $LR \in \mathfrak{G}$. The cosets of $\mathfrak{l}$ in $\mathfrak{L}$ are in 1:1 correspondence (given by the pairing) with the cosets of $\mathfrak{r}$ in $\mathfrak{R}$. For if L and L' are paired with R then LR, $(LR)^{-1} = L^{-1}R^{-1}$ and $L'R$ are in $\mathfrak{G}$. Therefore $L^{-1}R^{-1}L'R = L^{-1}L'R^{-1}R = L^{-1}L' \in \mathfrak{G}$. That is, $L^{-1}L' \in \mathfrak{l}$, and consequently L and L' are in the same coset of $\mathfrak{l}$. Conversely, if L and L' are in the same coset and if L is paired with R, that is, $LR \in \mathfrak{G}$, we have $L'L^{-1}LR = L'R \in \mathfrak{G}$. That is, L' is also paired with R. This correspondence is an isomorphism. For let L be in the coset corresponding to the coset containing R, that is, $LR \in \mathfrak{G}$, and let $L'R' \in \mathfrak{G}$. Then

$$LRL'R' = LL'RR' \in \mathfrak{G},$$

that is LL' is in the coset corresponding to the coset containing RR'.

This isomorphism

$$\mathfrak{L}/\mathfrak{l} \cong \mathfrak{R}/\mathfrak{r} \tag{3.1}$$

enables us to determine $\mathfrak{l}$ and $\mathfrak{r}$. For each of the $|\mathfrak{L}|$ elements of $\mathfrak{L}$ is paired with the $|\mathfrak{r}|$ elements of a coset of $\mathfrak{r}$ in $\mathfrak{R}$, and each of the $|\mathfrak{R}|$ elements of $\mathfrak{R}$ is paired with each of the $|\mathfrak{l}|$ elements of a coset of $\mathfrak{l}$ in $\mathfrak{L}$. These pairings give all the elements of $\mathfrak{G}$, but each element appears twice, for if $\{(a, 1)\}$ is paired with $\{(1, b)\}$ then $\{(-a, 1)\}$ is paired with $\{(1, -b)\}$ and $\{(a, 1)\}\ \{(1, b)\} = \{(-a, 1)\}\ \{(1, -b)\} = \{(a, b)\}$. That is, $|\mathfrak{G}| = \frac{1}{2}|\mathfrak{L}||\mathfrak{r}| = \frac{1}{2}|\mathfrak{R}||\mathfrak{l}|$. This gives $|\mathfrak{r}|$ and $|\mathfrak{l}|$ and consequently $\mathfrak{r}$ and $\mathfrak{l}$, since in all but two cases the normal subgroups of these orders are unique. We discuss these two cases separately.

If $\mathfrak{G} = 2[4]n$, n even, we have $\mathfrak{L} = \mathfrak{C}_{2n}$, $\mathfrak{R} = \langle 2, 2, n\rangle$ and $\mathfrak{l} = \mathfrak{C}_n$. In fact the $2n$ elements of $\mathfrak{L}$ are

$$\exp s\pi\mathbf{i}/n,\ s = 1, 2, \ldots, 2n.$$

(Strictly speaking, these are the elements of $\mathfrak{L}^*$. But $\mathfrak{L} \cong \mathfrak{L}^*$ and it is simpler to write $\pm a$ than $\{(\pm a, 1)\}$. The same applies to $\mathfrak{R}$ and $\mathfrak{R}^*$.) The $4n$ elements of $\mathfrak{R}$ are

$$\exp s\pi\mathbf{i}/n \text{ and } \mathbf{k}\exp s\pi\mathbf{i}/n,\ s = 1, 2, \ldots, 2n.$$

The possible choices of $\mathfrak{r}$, that is, of normal subgroups of index 2 in $\mathfrak{R}$, are $\mathfrak{C}_{2n}$ with elements $\exp s\pi\mathbf{i}/n$, $s = 1, 2, \ldots, 2n$, and $\langle 2, 2, n/2\rangle$ with elements

$$\exp 2s\pi\mathbf{i}/n \text{ and } \mathbf{k}\exp 2s\pi\mathbf{i}/n,\ s = 1, 2, \ldots, n.$$

We know $\mathfrak{G}$ contains the element $T = \{(\exp \pi\mathbf{i}/n,\ \exp -\pi\mathbf{i}/n)\}$. Since $\exp \pi\mathbf{i}/n$ is not in $\mathfrak{l}$, but in the other coset of $\mathfrak{l}$ in $\mathfrak{L}$ we know $\exp -\pi\mathbf{i}/n$ is not in $\mathfrak{r}$. But $\exp -\pi\mathbf{i}/n$ is in $\mathfrak{C}_{2n}$. Therefore $\mathfrak{r} = \langle 2, 2, n/2\rangle$.

In the case of groups $2[n]2$, n divisible by 4, where $\mathfrak{L} = \mathfrak{C}_4$, $\mathfrak{R} = \langle 2, 2, n/2\rangle$ and $\mathfrak{l} = \mathfrak{C}_2$, it can be verified in a similar manner that $\mathfrak{r} = \mathfrak{C}_n$ and not $\langle 2, 2, n/4\rangle$. For $T = \{(\mathbf{i}, -\mathbf{i})\}$ is in $\mathfrak{G}$ and $\mathbf{i}$ is not in $\mathfrak{l}$. But $-\mathbf{i}$ is in $\langle 2, 2, n/4\rangle$, so $\mathfrak{r} = \mathfrak{C}_n$. (In this case the elements of $\mathfrak{R}$ are $\exp 2s\pi\mathbf{j}/n$ and $\mathbf{i}\exp 2s\pi\mathbf{j}/n$, $s = 1, 2, \ldots, n$.)

Conversely, given the groups $\mathfrak{L}$, $\mathfrak{R}$, $\mathfrak{l}$, $\mathfrak{r}$ from our list we can always determine $\mathfrak{G}$ uniquely. To show this we need only show that if distinct isomorphisms 3.1 are chosen, the groups $\mathfrak{G}$ arising from the corresponding pairings of cosets are isomorphic. In most cases $\mathfrak{R}/\mathfrak{r}$ is $\mathfrak{C}_2$ or $\mathfrak{C}$, and there is thus only one isomorphism. However, there are cases where

$$\mathfrak{R}/\mathfrak{r} = \langle 2, 2, n\rangle/\mathfrak{C}_n \cong \mathfrak{C}_4$$

or

$$\mathfrak{R}/\mathfrak{r} = \langle 2, 3, 3\rangle/\langle 2, 2, 2\rangle \cong \mathfrak{C}_3.$$

There is only one non-trivial automorphism of $\mathfrak{C}_4$, and it is induced by an automorphism α of $\langle 2, 2, n\rangle$, namely by $\alpha b = (\mathbf{i})b(-\mathbf{i})$, where the elements b of $\langle 2, 2, n\rangle$ are again $\exp s\pi\mathbf{i}/n$ and $\mathbf{k}\exp s\pi\mathbf{i}/n$, $s = 1, 2, \ldots, 2n$. This establishes an isomorphism between the two possible groups $\mathfrak{G}$ by the correspondence $\{(a, b)\} : \{(a, \alpha b)\}$. Similarly, the only non-trivial automorphism of $\langle 2, 3, 3\rangle/\langle 2, 2, 2\rangle$ is induced by the automorphism

$$\beta b = \left(\frac{1 + \mathbf{i}}{\sqrt{2}}\right) b \left(\frac{1 - \mathbf{i}}{\sqrt{2}}\right)$$

of the group $\langle 2, 3, 3\rangle$ whose elements are $\pm\mathbf{1}$, $\pm\mathbf{i}$, $\pm\mathbf{j}$, $\pm\mathbf{k}$, $(\pm\mathbf{1} \pm \mathbf{i} \pm \mathbf{j} \pm \mathbf{k})$. That is, an isomorphism between the two groups obtained from the two possible pairings of cosets $\mathfrak{L}/\mathfrak{l}$ with those of $\langle 2, 3, 3\rangle/\langle 2, 2, 2\rangle$ is determined by the correspondence $\{(a, b)\} : \{(a, \beta b)\}$.

4. The relevant Goursat groups. Goursat **(5)** has shown that the finite groups of motions in elliptic 3-space can all be obtained in an analogous

fashion from pairings of corresponding cosets of isomorphic quotient groups of polyhedral or cyclic groups. Explicitly, consider a polyhedral or cyclic group $\mathfrak{L}'$ with a normal subgroup $\mathfrak{l}'$ and a polyhedral or cyclic group $\mathfrak{R}'$ with a normal subgroup $\mathfrak{r}'$ such that

$$\mathfrak{L}'/\mathfrak{l}' \cong \mathfrak{R}'/\mathfrak{r}'. \tag{4.1}$$

Then $|\mathfrak{L}'||\mathfrak{r}'| = |\mathfrak{R}'||\mathfrak{l}'|$ is the order of a group $\mathfrak{G}'$ whose elements are the pairs (a, b) for which b is an element of the coset of $\mathfrak{r}'$ in $\mathfrak{R}'$ which corresponds to the coset containing a in the isomorphism 4.1. The multiplication of elements of $\mathfrak{G}'$ is defined by $(a, b)\ (c, d) = (ac, bd)$, and $\mathfrak{G}'$ is a subgroup of $\mathfrak{L}' \times \mathfrak{R}'$.

In all but one of the cases which concern us the quotient groups 4.1 are either $\mathfrak{C}_2$ or $\mathfrak{C}$. Consequently the given construction for $\mathfrak{G}'$ is unambiguous. In the remaining case, where $\mathfrak{R}' = (2, 3, 3)$ and $\mathfrak{r}' = (2, 2, 2)$, there are two distinct ways of pairing the cosets of $\mathfrak{L}'/\mathfrak{l}'$ with those of $\mathfrak{R}'/\mathfrak{r}'$. But these two pairings again lead to isomorphic groups, for there is an automorphism of $(2, 3, 3)$ which induces the non-trivial automorphism of $(2, 3, 3)/(2, 2, 2)$ as follows. Let $(2, 3, 3)$ be the group of a regular tetrahedron, and let γ be a rotation by angle π about a line joining the midpoints of opposite edges of a cube whose vertices are those of the tetrahedron and its dual. Then the transformation $\gamma b \gamma^{-1}$ induces the non-trivial automorphism of $(2, 3, 3)/(2, 2, 2)$. Consequently an isomorphism between the two possible pair groups is given by the correspondence $(a, b) : (a, \gamma b \gamma^{-1})$.

5. The homomorphism from Shephard's groups to Goursat's. Let a group $\mathfrak{G}$ of our list be given by 3.1. Consider the natural homomorphism from the cyclic or binary polyhedral group $\mathfrak{L}$ to the corresponding cyclic or polyhedral group $\mathfrak{L}'$ obtained by identifying the elements $\pm a$ of $\mathfrak{L}$. Let a' be the image of $\pm a$ under this homomorphism. Similarly let b' be the image of $\pm b$ under the natural homomorphism of $\mathfrak{R}$ onto $\mathfrak{R}'$. Clearly this induces a homomorphism from $\mathfrak{G}$ onto some group $\mathfrak{G}'$ whose elements are of the form (a', b'). Thus $\mathfrak{G}'$ is one of Goursat's groups, for its elements are pairs from cyclic or polyhedral groups, and Goursat's list includes all such. We distinguish two cases. (i) The elements $\pm a$ are not in the same coset of $\mathfrak{l}$ in $\mathfrak{L}$, and the elements $\pm b$ are not in the same coset of $\mathfrak{r}$ in $\mathfrak{R}$. The only groups for which this occurs are $2[n]2$, n odd, and $2[4]n$, n odd, that is, the groups for which $\mathfrak{l}$ and $\mathfrak{r}$ are cyclic groups of odd order. (ii) The elements $\pm a$ are in the same coset of $\mathfrak{l}$ in $\mathfrak{L}$ and the elements $\pm b$ are in the same coset of $\mathfrak{r}$ in $\mathfrak{R}$. This is the situation for all other groups in our list.

Now in case (i) the homomorphism described from $\mathfrak{G}$ to $\mathfrak{G}'$ is actually an isomorphism. For the only elements of $\mathfrak{G}$ whose images are (a', b') are $\{(a, b)\}$ and $\{(-a, -b)\}$, since if $\{(a, b)\}$ is in $\mathfrak{G}$ there is no element $\{(a, -b)\}$ in $\mathfrak{G}$. But $\{(a, b)\} = \{(-a, -b)\}$, so the correspondence between the elements of $\mathfrak{G}'$ and those of $\mathfrak{G}$ is 1:1. This determines the order of $\mathfrak{G}'$, as well as $\mathfrak{L}'$ and $\mathfrak{R}'$, so that we can find $\mathfrak{l}'$ and $\mathfrak{r}'$ immediately. In fact, $\mathfrak{l}' = \mathfrak{l}$ and $\mathfrak{r}' = \mathfrak{r}$.

In case (ii) the homomorphism of $\mathfrak{G}$ to $\mathfrak{G}'$ is 2:1 since the distinct elements $\{(a, \pm b)\}$ of $\mathfrak{G}$ both have (a', b') as image. In particular the image of $(1, \pm b)$ is $(1', b')$, so if $b \in \mathfrak{r}$ then $b' \in \mathfrak{r}'$. That is, $\mathfrak{r}'$ is the image of $\mathfrak{r}$ under the homo-

TABLE I

Group $p_1[l]p_2$	Quaternion transformation corresponding to S^{-1} (§ 2)	$\mathfrak{L}/\mathfrak{l}$ (§ 3)	$\mathfrak{R}/\mathfrak{r}$ (§ 3)
$2[4]n$	$(\mathbf{i})q(-\mathbf{k})$	$\mathfrak{C}_{2n}/\mathfrak{C}_n$ (n even) $\mathfrak{C}_{4n}/\mathfrak{C}_n$ (n odd)	$\langle 2, 2, n\rangle/\langle 2, 2, n/2\rangle$ $\langle 2, 2, n\rangle/\mathfrak{C}_n$
$2[n]2$	$(\mathbf{i})q(-\mathbf{i}\exp 2\pi\mathbf{j}/n)$	$\mathfrak{C}_4/\mathfrak{C}_2$ (n even) $\mathfrak{C}_4/\mathfrak{C}$ (n odd)	$\langle 2, 2, n/2\rangle/\mathfrak{C}_n$ $\langle 2 ,2, n\rangle/\mathfrak{C}_n$
$3[6]2$	$\left(\frac{1}{2}+\mathbf{i}\frac{\sqrt{3}}{2}\right)q\left(\frac{1}{2}-\mathbf{i}\frac{1}{2}+\mathbf{j}\frac{\sqrt{2}}{4}-\mathbf{k}\frac{\sqrt{6}}{4}\right)$	$\mathfrak{C}_{12}/\mathfrak{C}_4$	$\langle 2, 3, 3\rangle/\langle 2, 2, 2\rangle$
$3[4]3$	$\left(\frac{1}{2}+\mathbf{i}\frac{\sqrt{3}}{2}\right)q\left(\frac{1}{2}-\mathbf{i}\frac{\sqrt{3}}{6}+\mathbf{j}\frac{\sqrt{6}}{6}-\mathbf{k}\frac{\sqrt{2}}{2}\right)$	$\mathfrak{C}_6/\mathfrak{C}_6$	$\langle 2, 3, 3\rangle/\langle 2, 3, 3\rangle$
$3[3]3$	$\left(\frac{1}{2}+\mathbf{i}\frac{\sqrt{3}}{2}\right)q\left(\frac{1}{2}+\mathbf{i}\frac{\sqrt{3}}{6}+\mathbf{j}\frac{\sqrt{6}}{6}-\mathbf{k}\frac{\sqrt{2}}{2}\right)$	$\mathfrak{C}_6/\mathfrak{C}_2$	$\langle 2, 3, 3\rangle/\langle 2, 2, 2\rangle$
$3[8]2$	$\left(\frac{1}{2}+\mathbf{i}\frac{\sqrt{3}}{2}\right)q\left(\frac{1}{2}-\mathbf{i}\frac{\sqrt{2}}{2}+\mathbf{j}\frac{1}{4}-\mathbf{k}\frac{\sqrt{3}}{4}\right)$	$\mathfrak{C}_{12}/\mathfrak{C}_6$	$\langle 2, 3, 4\rangle/\langle 2, 3, 3\rangle$
$4[6]2$	$\left(\frac{\sqrt{2}}{2}+\mathbf{i}\frac{\sqrt{2}}{2}\right)q\left(\frac{\sqrt{2}}{2}-\mathbf{i}\frac{1}{2}+\mathbf{j}\frac{\sqrt{2}}{4}-\mathbf{k}\frac{\sqrt{2}}{4}\right)$	$\mathfrak{C}_8/\mathfrak{C}_8$	$\langle 2, 3, 4\rangle/\langle 2, 3, 4\rangle$
$4[4]3$	$\left(\frac{\sqrt{2}}{2}+\mathbf{i}\frac{\sqrt{2}}{2}\right)q\left(\frac{\sqrt{2}}{2}-\mathbf{i}\frac{\sqrt{6}}{6}+\mathbf{j}\frac{\sqrt{6}}{6}-\mathbf{k}\frac{\sqrt{6}}{6}\right)$	$\mathfrak{C}_{24}/\mathfrak{C}_{12}$	$\langle 2, 3, 4\rangle/\langle 2, 3, 3\rangle$
$4[3]4$	$\left(\frac{\sqrt{2}}{2}+\mathbf{i}\frac{\sqrt{2}}{2}\right)q\left(\frac{\sqrt{2}}{2}+\mathbf{j}\frac{\sqrt{2}}{2}\right)$	$\mathfrak{C}_8/\mathfrak{C}_4$	$\langle 2, 3, 4\rangle/\langle 2, 3, 3\rangle$
$3[10]2$	$\left(\frac{1}{2}+\mathbf{i}\frac{\sqrt{3}}{2}\right)q\left(\frac{1}{2}-\mathbf{i}\frac{\tau^*}{2}+\mathbf{j}\frac{1}{2\tau}-\mathbf{k}\frac{\sqrt{3}}{4\tau}\right)$	$\mathfrak{C}_{12}/\mathfrak{C}_{12}$	$\langle 2, 3, 5\rangle/\langle 2, 3, 5\rangle$
$5[6]2$	$\left(\frac{\tau}{2}+\mathbf{i}\frac{\sigma}{2}\right)^{\dagger}q\left(\frac{\tau}{2}-\mathbf{i}\frac{1}{2}+\mathbf{j}\frac{1}{4}-\mathbf{k}\frac{\sigma^3\sqrt{5}}{20}\right)$	$\mathfrak{C}_{20}/\mathfrak{C}_{20}$	$\langle 2, 3, 5\rangle/\langle 2, 3, 5\rangle$
$5[4]3$	$\left(\frac{\tau}{2}+\mathbf{i}\frac{\sigma}{2}\right)q\left(\frac{\tau}{2}-\mathbf{i}\frac{\tau\sqrt{3}}{6}+\mathbf{j}\frac{\sqrt{3}}{6}+\mathbf{k}\frac{\sigma^3\sqrt{15}}{30}\right)$	$\mathfrak{C}_{30}/\mathfrak{C}_{20}$	$\langle 2, 3, 5\rangle/\langle 2, 3, 5\rangle$
$3[5]3$	$\left(\frac{1}{2}+\mathbf{i}\frac{\sqrt{3}}{2}\right)q\left(\frac{1}{2}-\mathbf{i}\frac{\sqrt{15}}{6}+\mathbf{j}\frac{\sqrt{3}}{6}-\mathbf{k}\frac{1}{2}\right)$	$\mathfrak{C}_6/\mathfrak{C}_6$	$\langle 2, 3, 5\rangle/\langle 2, 3, 5\rangle$
$5[3]5$	$\left(\frac{\tau}{2}+\mathbf{i}\frac{\sigma}{2}\right)q\left(\frac{\tau}{2}-\mathbf{i}\frac{\sigma\sqrt{5}}{10}+\mathbf{j}\frac{1}{2\sigma}-\mathbf{k}\frac{1}{2\tau}\right)$	$\mathfrak{C}_{10}/\mathfrak{C}_{10}$	$\langle 2, 3, 5\rangle/\langle 2, 3, 5\rangle$

$^*\tau = 2\cos\pi/5 = (1+\sqrt{5})/2$. $\quad$ $^\dagger\sigma = 2\sin\pi/5 = (3-\tau)^{\frac{1}{2}}$.

morphism taking $\pm b$ to b'. This determines the group $\mathfrak{r}'$. Order considerations alone determine the cyclic group $\mathfrak{l}'$ for which $2|\mathfrak{l}'| = |\mathfrak{l}|$.

The groups $\mathfrak{L}$, $\mathfrak{R}$, $\mathfrak{l}$, $\mathfrak{r}$ appear in Table I. These tabulations then readily yield the Goursat groups in the form 4.1. Except for the cases $2[4]n$ and $2[n]2$ the results are summarized in the Theorem, for which the following notation is convenient:

Let $p_1[t]p_2$ be the group generated by reflections S^{-1} and T, having the defining relations

$$(S^{-1})^{p_1} = T^{p_2} = E, S^{-1}TS^{-1}\ldots = TS^{-1}T\ldots \text{ (t factors on each side).}$$

The centre of this group is the cyclic group $\mathfrak{Z}$ generated by $(S^{-1}T)^{t/2}$ if t is even or by $(S^{-1}T)^t$ if t is odd. The period of $(S^{-1}T)^{t/2}$ is

$$2p_1p_2/k, \text{ where } 2k = 2p_1p_2 + p_1t + p_2t - p_1p_2t$$

(**4**, pp. 76, 77, 79). The quotient group $p_1[t]p_2/\mathfrak{Z}$ is a polyhedral group $(2, \mu, \nu)$. (**7**, p. 84).

THEOREM. *The group $p_1[t]p_2$ ($p_1 \neq 2$) with centre $\mathfrak{Z}$ is 2:1 homomorphic to the group of motions in elliptic 3-space defined by the isomorphism*

$$\mathfrak{L}'/\mathfrak{l}' \cong \mathfrak{R}'/\mathfrak{r}',$$

where

(a) *$\mathfrak{L}'$ and $\mathfrak{l}'$ are cyclic groups.*
(b) $|\mathfrak{L}'| = \text{l.c.m.}\{p_1, p_2\}$.
(c) $2|\mathfrak{l}'| = |\mathfrak{Z}|$.
(d) $\mathfrak{R}' = p_1[t]p_2/\mathfrak{Z}$.
(e) *$\mathfrak{r}'$ is the unique normal subgroup of $\mathfrak{R}'$ such that $|\mathfrak{L}'||\mathfrak{r}'| = |\mathfrak{R}'||\mathfrak{l}'|$.*

((a), (b), and (d) also hold in case $p_1 = 2$.)

REFERENCES

1. H. S. M. Coxeter, *The binary polyhedral groups and other generalisations of the quaternion group*, Duke Math. J., *7* (1940), 367–379.
2. ——— *Quaternions and reflections*, Amer. Math. Monthly, *53* (1946), 136–146.
3. ——— *Regular polytopes* (London, 1948).
4. H. S. M. Coxeter and W. O. J. Moser, *Generators and relations for discrete groups* (Berlin, 1957).
5. E. Goursat, *Sur les substitutions orthogonales et les divisions régulières de l'espace*, Ann. Sci. Ecole Norm. Sup. (3), *6* (1889), 9–102.
6. H. Seifert and W. Threlfall, *Topologische Untersuchungen der Diskontinuitätsbereiche endlicher Bewegungsgruppen des dreidimensionalen sphärischen Raumes*, Math. Ann., *104* (1931), 1–70.
7. G. C. Shephard, *Regular complex polygons*, Proc. London Math. Soc. (3), *2* (1952), 82–97.
8. G. C. Shephard and J. A. Todd, *Finite unitary reflection groups*, Can. J. Math., *6* (1954), 274–304.

University College
Ibadan, Nigeria.

DECOMPOSITION OF FINITE GRAPHS INTO OPEN CHAINS

C. ST. J. A. NASH-WILLIAMS

1. Introduction. If m, n are integers, "$m \equiv n$" will mean "$m \equiv n$ (mod 2)." The cardinal number of a set A will be denoted by $|A|$. The set whose elements are $a_1, a_2, \ldots, a_n$ will be denoted by $\{a_1, a_2, \ldots, a_n\}$. The empty set will be denoted by $\emptyset$. If A, B, C are sets, $A - B$ will denote the set of those elements of A which do not belong to B, and $A - B - C$ will denote $(A - B) - C$. The expression $\sum_{\xi \in A} f(\xi)$ will be denoted by $f \,.\, A$. The statements "$f = g$ on A," "$f \equiv g$ on A" will mean that $f(\xi) = g(\xi)$ or $f(\xi) \equiv g(\xi)$ respectively for every $\xi \in A$.

An ***unoriented graph*** U consists, for the purposes of this paper, of two disjoint finite sets $V(U)$, $E(U)$, together with a relationship whereby with each $\lambda \in E(U)$ is associated an unordered pair of (not necessarily distinct) elements of $V(U)$ which λ is said to ***join***. An ***oriented graph*** is a triple $N = (U, t, h)$, where U is an unoriented graph and t, h are mappings of $E(U)$ into $V(U)$ such that each $\lambda \in E(U)$ joins λt to λh. We write $V(U) = V(N)$, $E(U) = E(N)$ and call λt, λh the ***tail*** and ***head*** of λ respectively. Either an unoriented or an oriented graph may be referred to as a ***graph***. Throughout this paper, U will denote an unoriented graph, N will denote an oriented graph, and G may denote either. The elements of $V(G)$ and $E(G)$ are called ***vertices*** and ***edges*** of G respectively. A ***subgraph*** of U is an unoriented graph H such that $V(H) \subset V(U)$, $E(H) \subset E(U)$ and each edge of H joins the same vertices in H as in U. A ***subgraph*** of $N = (U, t, h)$ is an oriented graph (U_1, t_1, h_1) such that U_1 is a subgraph of U and t_1, h_1 are the restrictions of t, h respectively to $E(U_1)$. An ***orientation*** of U is an oriented graph of the form (U, t, h). A vertex ξ and edge λ of G are ***incident*** if ξ is one or both of the vertices joined by λ. The ***order***, ord G, of G is $|V(G) \cup E(G)|$. G is ***empty*** if $V(G) = E(G) = \emptyset$. The ***degree*** $d(\xi)$ of a vertex ξ of a graph is $2a(\xi) + b(\xi)$, where $a(\xi)$ is the number of edges joining ξ to itself and $b(\xi)$ is the number joining ξ to other vertices. A vertex is ***even*** or ***odd*** according as its degree is even or odd respectively. G is ***Eulerian*** if its vertices are all even. A collection of subgraphs of G are ***disjoint*** (***edge-disjoint***) if no two of them have a vertex (edge) in common. The ***union*** of the subgraphs $H_1, H_2, \ldots, H_n$ of G is the subgraph H of G such that

$$V(H) = \bigcup_{i=1}^{n} V(H_i), \qquad E(H) = \bigcup_{i=1}^{n} E(H_i).$$

A ***decomposition*** of G is a set of edge-disjoint subgraphs of G whose union is G. G is ***connected*** if it is not the union of two disjoint non-empty subgraphs. The

Received December 7, 1959.

components of a non-empty graph are its maximal connected subgraphs. (An empty graph is deemed to have 0 components.) A ***chain-sequence*** of G is a finite sequence

$$\xi_0, \lambda_1, \xi_1, \lambda_2, \xi_2, \lambda_3, \ldots, \lambda_n, \xi_n \qquad (n > 0)$$

in which the ξ_i are vertices of G, the λ_i are *distinct* edges of G and λ_i joins ξ_{i-1} to ξ_i for $i = 1, 2, \ldots, n$. If G is an oriented graph, this chain-sequence is ***forwards-directed*** if

$$\lambda_i t = \xi_{i-1}, \lambda_i h = \xi_i \qquad (i = 1, 2, \ldots, n)$$

and ***backwards-directed*** if

$$\lambda_i h = \xi_{i-1}, \lambda_i t = \xi_i \qquad (i = 1, 2, \ldots, n).$$

A finite sequence is ***closed*** or ***open*** according as its first and last terms are the same or different respectively. If s is a chain-sequence of G, the subgraph of G formed by those vertices which appear at least once and those edges which appear exactly once in s will be said to be ***derived*** from s. A subgraph of G is an ***open chain*** of G if it is derivable from an open chain-sequence of G. If ξ, η are the first and last terms of an open chain-sequence s of G and C is the open chain derived from s, then clearly ξ, η are odd in C and every other vertex of C is even in C. It follows that an open chain has precisely two odd vertices which are the end-terms of every chain-sequence from which it is derivable; these are called the ***end-vertices*** of the open chain. If S, T are subsets of $V(G)$, $\bar{S}$ will denote $V(G) - S$, $S \circ T$ will denote the set of those edges of G which join elements of S to elements of T, and $S\delta$ will denote $S \circ \bar{S}$. A subgraph of G is an ***ST-chain*** if it is derivable from a chain-sequence of G whose first and last terms belong to S, T respectively. A ***cincture*** of G is a subset of $E(G)$ which is of the form $S\delta$ for some subset S of $V(G)$. If $\xi \in V(N)$, an edge λ is an ***exit*** of ξ if $\lambda t = \xi$ and an ***entry*** of ξ if $\lambda h = \xi$. The number of exits [entries] of ξ will be denoted by $x(\xi)$ [$e(\xi)$]. The ***flux out of*** ξ, denoted by $f(\xi)$, is $x(\xi) - e(\xi)$. N is ***quasi-symmetrical*** if $x = e$ on $V(N)$. A ***route-sequence*** of N is a chain-sequence of N which is either forwards- or backwards-directed. A subgraph of N is a ***route*** (***closed route***, ***open route***) of N if it is derivable from a route-sequence (closed route-sequence, open route-sequence) of N.

When, to avoid ambiguity, it is necessary to specify the graph relative to which a graph-theoretical symbol is defined, the letter denoting the graph will be attached to the symbol in some convenient way. For example, if ξ is a common vertex of two oriented graphs M and N, $d_M(\xi)$ will denote the degree of ξ in M. We shall, however, make the convention that, in any context in which an oriented graph denoted by the letter N is under consideration, all graph-theoretical symbols relate to N unless the contrary is indicated; for example, $d(\xi)$ would mean $d_N(\xi)$ in the situation instanced above.

Let s be a forwards-directed route-sequence of N, R be the route derived from s and ξ, η be the first and last terms of s respectively. Then clearly R is

quasi-symmetrical if $\xi = \eta$ and $f_R(\xi) = 1, f_R(\eta) = -1$ and $f_R = 0$ on $V(R) - \{\xi, \eta\}$ if $\xi \neq \eta$. It follows that a closed route cannot also be an open route and that an open route R has uniquely determined vertices ξ, η such that $f_R(\xi) = 1$, $f_R(\eta) = -1$ and ξ, η are the first and last terms respectively of every forwards-directed route-sequence from which R is derivable; we call ξ, η the *tail* and *head* respectively of R.

By a *G-function*, we shall mean a non-negative integer-valued function defined on the vertices of G. A G-function g is *congruential* if $g \equiv d$ on $V(G)$. If g is a G-function and $\xi \in S \subset V(G)$, $F_g(\xi; S)$ will denote

$$-g(\xi) + g\,.\,(S - \{\xi\}) + |S\delta|.$$

We shall call g *tolerable* if $F_g(\xi; S) \geqslant 0$ for every pair ξ, S such that $\xi \in S \subset V(G)$. A subset S of $V(G)$ is *g-critical* if $F_g(\xi; S) = 0$ for some $\xi \in S$. A cincture C of G is *g-critical* if $C = S\delta$ for some g-critical subset S of $V(G)$. A *g-chain-factor* of G is a set Φ of edge-disjoint open chains of G such that each vertex ξ of G is an end-vertex of exactly $g(\xi)$ elements of Φ. A *g-decomposition* of G is a g-chain-factor of G which is a decomposition of G.

Let u, v be N-functions. Then a *(u, v)-route-factor* of N is a set Φ of edge-disjoint open routes of N such that each vertex ξ of N is the tail of exactly $u(\xi)$ and head of exactly $v(\xi)$ elements of Φ. A *(u, v)-decomposition* of N is a (u, v)-route-factor of N which is a decomposition of N.

The object of this paper is to prove the following two parallel results:

THEOREM 1. *Let g be a U-function. Then U has a g-decomposition if and only if g is tolerable and congruential and $g\,.\,V(H) > 0$ for each component H of U.*

THEOREM 2. *Let u, v be N-functions. Then N has a (u, v)-decomposition if and only if $u + v$ is tolerable, $u - v = f$ on $V(N)$ and $(u + v)\,.\,V(H) > 0$ for each component H of N.*

Our procedure will be to prove Theorem 2 and deduce Theorem 1 from it. Certain generalizations of the theorems will be mentioned at the end of the paper.

2. Proof of Theorem 2.

LEMMA 1. *If G has a g-chain-factor, g is tolerable.*

Proof. Let Φ be a g-chain-factor of G. For any pair of disjoint subsets S, T of $V(G)$, let $S*T$ denote the number of ST-chains in Φ. Then, if $\xi \in S \subset V(G)$,

$$g(\xi) = (\{\xi\}*\bar{S}) + \sum_{\eta \in S - \{\xi\}} (\{\xi\}*\{\eta\}).$$

But $\{\xi\}*\{\eta\} \leqslant g(\eta)$ for every $\eta \in S - \{\xi\}$; and $\{\xi\}*\bar{S} \leqslant |S\delta|$ since $\xi \in S$ and so each $\{\xi\}\bar{S}$-chain must include an element of $S\delta$. Hence $g(\xi) \leqslant g\,.\,(S - \{\xi\}) + |S\delta|$; and the lemma is proved.

LEMMA 2. *If A, B are disjoint subsets of $V(G)$, $|(A \cup B)\delta| + |A\delta| \geqslant |B\delta|$.*

Proof. If $V(G) - (A \cup B) = C$, the above inequality follows from the relations

$$|A\delta| = |A \circ B| + |C \circ A|, |B\delta| = |B \circ C| + |A \circ B|, |(A \cup B)\delta| = |C \circ A| + |B \circ C|.$$

LEMMA 3. *If $S \subset V(G)$, $|S\delta| \equiv d \,.\, S$.*

Proof. An edge contributes 2, 1, or 0 to $d \,.\, S$ according as it belongs to $S \circ S$, $S\delta$ or $\bar{S} \circ \bar{S}$ respectively.

COROLLARY 3A. *If g is a congruential G-function and $\xi \in S \subset V(G)$, $F_g(\xi; S)$ is even.*

COROLLARY 3B. (=(**1**, chapter II, Theorem **3**)). *The number of odd vertices of a graph is even.*

Proof. Take $S = V(G)$ in Lemma 3.

Definition. Let λ, μ be distinct edges of N such that $\lambda h = \mu t = \xi$. Then the oriented graph M obtained from N by *fusion of* λ *and* μ *at* ξ is defined by the rules:

(i) $V(M) = V(N)$, $E(M) = [E(N) - \{\lambda, \mu\}] \cup \{\nu\}$, where ν is a newly added edge and is not an element of the set $V(N) \cup E(N)$;

(ii) $\nu t_M = \lambda t$, $\nu h_M = \mu h$;

(iii) $\kappa t_M = \kappa t$, $\kappa h_M = \kappa h$ for every $\kappa \in E(N) - \{\lambda, \mu\}$.

LEMMA 4. *If, in the circumstances of the above definition, g is a tolerable congruential N-function and no g-critical cincture of N includes both λ and μ, then g is tolerable in M.*

Proof. Let $\xi \in S \subset V(M)$ $(= V(N))$. If λ, μ do not both belong to $S\delta$, then $|S\delta_M| = |S\delta|$ and so ${}_MF_g(\xi; S) = F_g(\xi; S) > 0$. If λ, μ both belong to $S\delta$, then (i) $|S\delta_M| = |S\delta| - 2$, whence ${}_MF_g(\xi; S) = F_g(\xi; S) - 2$, and (ii) $S\delta$ must not be g-critical, whence, by the tolerability of g and Corollary 3A, $F_g(\xi; S) > 2$. Hence ${}_MF_g(\xi; S) > 0$.

Definitions. If $S \subset V(N)$, S^* will denote the subgraph of N defined by $V(S^*) = S$, $E(S^*) = S \circ S$, and N_S will denote the oriented graph M defined as follows.

(i) $V(M) = \bar{S} \cup \{S'\}$, $E(M) = \bar{S} \circ V(N)$, where S' $[\notin V(N) \cup E(N)]$ is a newly introduced vertex.

(ii) Write $\phi(\xi) = \xi$ if $\xi \in \bar{S}$ and $\phi(\xi) = S'$ if $\xi \in S$. Then $\lambda t_M = \phi(\lambda t)$, $\lambda h_M = \phi(\lambda h)$ for every $\lambda \in E(M)$.

Thus N_S is obtained from N by contracting the subgraph S^* to a single vertex S'.

LEMMA 5. *Let g be a tolerable N-function and C be a g-critical subset of $V(N)$. If $g(C')$, $g(\bar{C}')$ are both defined to be $|C\delta|$, then g is tolerable in N_C and $N_{\bar{C}}$.*

Proof. Write $N_{\bar{C}} = H$, $N_C = K$. Since C is critical,

$$g(\xi) = g \,.\, (C - \{\xi\}) + |C\delta| \tag{1}$$

for some $\xi \in C$. Since $g(C') = g(\bar{C}') = |C\delta|$, (1) can be rewritten in each of the forms

$$(1') \qquad g(\xi) = g \,.\, [V(H) - \{\xi\}],$$

$$(1'') \qquad g(C') = g(\xi) - g \,.\, (C - \{\xi\}).$$

LEMMA 5A[1]. *If* $S \subset V(H) - \{\xi\}$, $g \,.\, S < |S\delta_H|$.

Proof. Since $F_g(\xi; C - S) \geqslant 0$,

$$(2) \qquad g(\xi) - g \,.\, (C - S - \{\xi\}) \leqslant |(C - S)\delta|.$$

If $\bar{C}' \notin S$,

$$|S\delta_H| = |S\delta| \geqslant |(C - S)\delta| - |C\delta| \geqslant g(\xi) - g \,.\, (C - S - \{\xi\}) - |C\delta| = g \,.\, S$$

by Lemma 2, (2) and (1). If $\bar{C}' \in S$,

$$|S\delta_H| = |(C - S)\delta| \geqslant g(\xi) - g \,.\, (C - S - \{\xi\}) = g \,.\, S$$

by (2) and (1′).

Suppose that $Y \subset V(H)$. Let $V(H) - Y = W$. If $\xi \notin Y$, then, for every $\eta \in Y$,

$$_HF_g(\eta; Y) \geqslant |Y\delta_H| - g(\eta) \geqslant |Y\delta_H| - g \,.\, Y \geqslant 0$$

by Lemma 5A. If $\xi \in Y$, then by (1′),

$$_HF_g(\xi; Y) = |Y\delta_H| - g \,.\, W = |W\delta_H| - g \,.\, W \geqslant 0$$

by Lemma 5A, and, for every $\eta \in Y - \{\xi\}$,

$$_HF_g(\eta; Y) \geqslant g \,.\, (Y - \{\eta\}) - g(\eta) \geqslant 0$$

by (1′). Hence g is tolerable in H.

Suppose that $Z \subset V(K)$. If $C' \notin Z$, then $Z\delta_K = Z\delta$ and so $_KF_g(\eta; Z) = F_g(\eta; Z) \geqslant 0$ for every $\eta \in Z$. If $C' \in Z$, then

$$(3) \qquad Z\delta_K = \bar{Z}\delta$$

where $\bar{Z} = (Z - \{C'\}) \cup C$. By (1″) and (3), $_KF_g(C'; Z) = F_g(\xi; \bar{Z}) \geqslant 0$; and, by (3) and Lemma 2,

$$g(C') + |Z\delta_K| = |C\delta| + |\bar{Z}\delta| \geqslant |(Z - \{C'\})\delta|,$$

whence $_KF_g(\eta; Z) \geqslant F_g(\eta; Z - \{C'\}) \geqslant 0$ for every $\eta \in Z - \{C'\}$. Hence g is tolerable in K.

Definitions. An edge λ of N is a *loop* if $\lambda t = \lambda h$. If g is an N-function, a vertex ξ is *g-critical* if the set $\{\xi\}$ is g-critical, that is, if $g(\xi) = |\{\xi\}\delta|$, and is *g-safe* if $F_g(\xi; \{\xi\}) > 0$, that is, if $g(\xi) < |\{\xi\}\delta|$. A *one-edge-route* is a route which has exactly one edge. If $S \subset V(N)$, an edge λ is an *exit* of S if $\lambda t \in S$, $\lambda h \in \bar{S}$, and is an *entry* of S if $\lambda h \in S$, $\lambda t \in \bar{S}$. If $A \subset E(N)$, $N - A$ will denote the

[1]We give the names Lemma nA, Lemma nB to lemmas which themselves form part of the proof of Lemma n.

subgraph of N defined by the relations $V(N - A) = V(N)$, $E(N - A) = E(N) - A$.

LEMMA 6. *If u and v are N-functions such that $u - v = f$ on $V(N)$ and $u + v$ is tolerable, then N has a (u, v)-route-factor.*

Proof. Since Lemma 6 is trivially true for an oriented graph of order 0, it may be proved by induction on ord N. We shall therefore make the inductive hypothesis that Lemma 6 is true for all oriented graphs of lower order than N. Let $u + v = g$. If N has a loop λ, then λ belongs to no cincture. Therefore g, being tolerable in N, is tolerable in $N - \{\lambda\}$. It is also clear that $f_{N-\{\lambda\}} = f = u - v$ on $V(N)$. Therefore, by the inductive hypothesis, $N - \{\lambda\}$ has a (u, v)-route-factor, and hence so has N. We shall therefore henceforward assume that N is loopless. We shall consider separately the following two cases: (I) $V(N)$ has a g-critical subset C such that $|C| > 2$ and $|\bar{C}| > 2$; (II) $V(N)$ has no such subset.

Proof for Case I. Let the exits of C be $\lambda_1, \lambda_2, \ldots, \lambda_p$ and its entries be $\lambda_{p+1}, \lambda_{p+2}, \ldots, \lambda_r$. If we write $N_C = K$, $u(C') = p$, $v(C') = r - p$ and $g(C') = |C\delta|$, then u, v, and g are defined on all vertices of K and $g = u + v$ on $V(K)$. By Lemma 5, g is tolerable in K. It is clear that $u(C') - v(C') = f_K(C')$ and that $f_K = f = u - v$ on $\bar{C}$; hence $u - v = f_K$ on $V(K)$. Since $|C| > 2$, ord $K <$ ord N. Therefore, by the inductive hypothesis, K has a (u, v)-route-factor Φ. Since $u(C') + v(C') = r$ and $\lambda_1, \lambda_2, \ldots, \lambda_r$ are the only edges incident with C' in K, it is clear that $\lambda_1, \lambda_2, \ldots, \lambda_r$ must be distributed in a one-to-one fashion amongst the r elements of Φ which have C' as an end-vertex; let R_i be that element of Φ which includes λ_i among its edges. Then clearly R_i is derivable from a route-sequence of the form C', λ_i, s_i, where s_i is a route-sequence of $\bar{C}^*$. Clearly C', λ_i, s_i and hence also s_i must be forwards- or backwards-directed according as C' is the tail or head respectively of λ_i in K, that is, according as $i < p$ or $i > p$ respectively. Moreover, if $\Phi - \{R_1, R_2, \ldots, R_r\} = \Delta$, then, since the λ_i are the only edges incident with C' in K and $\lambda_i \in E(R_i)$ $(i = 1, 2, \ldots, r)$, it follows that each element of Δ is a route of $\bar{C}^*$.

If we write $u(\bar{C}') = r - p$, $v(\bar{C}') = p$, an argument similar to that of the preceding paragraph, but using the hypothesis that $|\bar{C}| > 2$ and the assertion concerning $N_{\bar{C}}$ in Lemma 5, shows that $N_{\bar{C}}$ has a (u, v)-route-factor $\bar{\Delta} \cup \{\bar{R}_1, \bar{R}_2, \ldots, \bar{R}_r\}$ such that the elements of $\bar{\Delta}$ are routes of C^* and, for $i = 1, 2, \ldots, r$, $\bar{R}_i$ is derivable from a route-sequence of the form $\bar{s}_i, \lambda_i, \bar{C}'$, where $\bar{s}_i$ is a route-sequence of C^* and is forwards- or backwards-directed according as $i < p$ or $i > p$ respectively. It is now not difficult to see that, if S_i is the route derived from the route-sequence $\bar{s}_i, \lambda_i, s_i$, then $\Delta \cup \bar{\Delta} \cup \{S_1, S_2, \ldots, S_r\}$ is a (u, v)-route-factor of N.

Proof for Case II.

LEMMA 6A. *A vertex ξ of N is g-critical if $V(N) - \{\xi\}$ is g-critical.*

Proof. If $V(N) - \{\xi\}$ is g-critical,

$$-g(\eta) + g\,.\,(V(N) - \{\xi, \eta\}) + |\{\xi\}\delta| = 0$$

for some $\eta \in V(N) - \{\xi\}$. But

$$-g(\eta) + g\,.\,(V(N) - \{\xi, \eta\}) + g(\xi) = F_g(\eta; V(N)) \geqslant 0.$$

Therefore $g(\xi) \geqslant |\{\xi\}\delta|$, that is, $F_g(\xi; \{\xi\}) \leqslant 0$. Hence, since g is tolerable, $F_g(\xi; \{\xi\}) = 0$ and so ξ is g-critical.

Corollary 6AA. *In Case* II, *every non-empty g-critical cincture is of the form* $\{\xi\}\delta$ *for some g-critical vertex* ξ.

If ξ is a g-critical vertex, $g(\xi) = |\{\xi\}\delta|$, that is, since N is loopless, $u(\xi) + v(\xi) = x(\xi) + e(\xi)$. But, by hypothesis, $u(\xi) - v(\xi) = f(\xi) = x(\xi) - e(\xi)$. Hence $u(\xi) = x(\xi)$ and $v(\xi) = e(\xi)$. Hence, since N is loopless, the one-edge-routes in N constitute a (u, v)-route-factor of N if every vertex of N is g-critical. We may therefore assume that N has a g-safe vertex σ. Since σ is g-safe,

$$|\{\sigma\}\delta| > g(\sigma) \geqslant |u(\sigma) - v(\sigma)| = |f(\sigma)|$$

by hypothesis. Therefore

$$x(\sigma) > 0, \qquad e(\sigma) > 0. \tag{4}$$

Lemma 6B. *The vertex* σ *has an entry* λ *and an exit* μ *such that no g-critical cincture includes both* λ *and* μ.

Proof. (Throughout this proof, the reader should bear in mind that N is assumed to be loopless.) If σ is adjacent to two or more other vertices, it is easily seen from (4) that σ has an entry λ and an exit μ which join it to different vertices; since σ is g-safe and is the only vertex incident with both λ and μ, Corollary 6AA shows that no g-critical cincture includes both λ and μ. We may therefore assume that σ is adjacent to at most one, and hence, by (4), to exactly one other vertex; let this vertex be τ. Since σ is adjacent only to τ, $|\{\sigma, \tau\}\delta| = |\{\tau\}\delta| - |\{\sigma\}\delta|$. Therefore

$$-\,g(\tau) + g(\sigma) + |\{\tau\}\delta| - |\{\sigma\}\delta| = F_g(\tau; \{\sigma, \tau\}) \geqslant 0.$$

But $|\{\sigma\}\delta| > g(\sigma)$ since σ is g-safe. Therefore $|\{\tau\}\delta| > g(\tau)$. Hence τ is also g-safe. But, by (4), we can select an entry λ and an exit μ of σ. Since λ, μ must both join σ, τ, which are both g-safe, Corollary 6AA again implies the required result.

Since

$$g = u + v \equiv u - v = f = x - e \equiv x + e = d$$

on $V(N)$, g is congruential in N. Therefore, by Lemmas 6B and 4, g is tolerable in the oriented graph (M, say) obtained from N by fusion of λ and μ at σ. It is also clear that $f_M = f = u - v$ on $V(N) = V(M)$ and that ord M = ord N − 1. Therefore, by the inductive hypothesis, M has a (u, v)-route-factor, and it is easily seen that this is converted into a (u, v)-route-factor of N when we reverse the fusion of λ and μ at σ.

LEMMA 7. *If N has a decomposition of the form $\Phi \cup \Theta$, where Φ is a (u, v)-route-factor of N and Θ is a set of closed routes each of which has a vertex in common with some element of Φ, then N has a (u, v)-decomposition.*

Proof. Let $\Phi = \{R_1, R_2, \ldots, R_r\}$, and let $\Theta = \Theta_1 \cup \Theta_2 \cup \ldots \cup \Theta_r$, where the Θ_i are disjoint and each element of Θ_i has a vertex in common with R_i. If S_i is the union of R_i and the elements of Θ_i, it is easily seen that S_i is an open route with the same head and tail as R_i. Hence $\{S_1, S_2, \ldots, S_r\}$ is a (u, v)-decomposition of N.

Proof of Theorem 2. The necessity of the first condition follows from Lemma 1, and the necessity of the other two is obvious. Conversely, suppose that these three conditions are satisfied. Then, by Lemma 6, N has a (u, v)-route-factor Φ. If T is the union of the elements of Φ, then clearly $f_T = u - v$ on $V(T)$ and $u = v = 0$ on $V(N) - V(T)$. But $f = u - v$ on $V(N)$ by hypothesis. Therefore $N - E(T)$ is quasi-symmetrical. Therefore, by (**1**, chapter II, Theorem 7), every component of $N - E(T)$ is a closed route. Moreover, since $(u + v) \,.\, V(H) > 0$ for each component H of N, each component of N contains an element of Φ and hence each component of $N - E(T)$ has a vertex in common with an element of Φ. Therefore, by Lemma 7 (with Θ taken to be the set of components of $N - E(T)$), N has a (u, v)-decomposition.

3. Proof of Theorem 1.

LEMMA 8. *Every unoriented graph has an orientation in which $f(\xi) = 0$ for each even vertex ξ and $f(\xi) = \pm 1$ for each odd vertex ξ.*

Proof. Let U be a given unoriented graph. By Corollary 3B, the number of odd vertices of U is even; let it be $2r$. Then U can be converted into an Eulerian unoriented graph H by the addition of r new edges joining its odd vertices in pairs.[2] H, being Eulerian, has by (**1**, p. 30, ll. 4-9), a quasi-symmetrical orientation, and this clearly induces in U an orientation of the required type.

Proof of Theorem 1. The necessity of the condition that g be tolerable follows from Lemma 1, and the necessity of the remaining conditions is obvious. Conversely, let the conditions of Theorem 1 be satisfied, and let N be an orientation of U satisfying the condition of Lemma 8. Write $u = \frac{1}{2}(g + f)$, $v = \frac{1}{2}(g - f)$, where f denotes flux in N. Then, by Theorem 2, N has a (u, v)-decomposition, and hence U has a g-decomposition.

4. Generalizations.

Definitions. A *semi-oriented graph* is a quintuple $S = (U, \mathfrak{x}, \mathfrak{e}, p, q)$ such that U is an unoriented graph, $\mathfrak{x}, \mathfrak{e}$ are disjoint sets and p, q are mappings of $\mathfrak{x} \cup \mathfrak{e}$ into $V(U)$, $E(U)$ respectively, subject to the condition that each edge λ of U is the image under q of exactly two elements of $\mathfrak{x} \cup \mathfrak{e}$ and that, if these elements are ϵ, ϵ', then λ joins ϵp to $\epsilon' p$ in U. Vertices and edges of U are

[2]This procedure is suggested by the proof of (**1**, chapter II, Theorem 4).

called *vertices* and *edges* of S respectively, and elements of $\mathfrak{r} \cup \mathfrak{e}$ are called *hinges* of S. A vertex ξ (edge λ) of U is *incident* with a hinge ϵ if $\epsilon p = \xi$ ($\epsilon q = \lambda$). Two hinges are *opposed* if one of them belongs to $\mathfrak{r}$ and the other to $\mathfrak{e}$. If $\xi \in V(U)$, $f(\xi)$ will denote $|\mathfrak{s} \cap \mathfrak{r}| - |\mathfrak{s} \cap \mathfrak{e}|$, where $\mathfrak{s}$ is the set of those hinges of S which are incident with ξ. An *open route-sequence* of S is a finite sequence

$$(5) \qquad \xi_0, \epsilon_1, \lambda_1, \bar{\epsilon}_1, \xi_1, \epsilon_2, \lambda_2, \bar{\epsilon}_2, \xi_2, \epsilon_3, \ldots, \lambda_n, \bar{\epsilon}_n, \xi_n$$

such that $\xi_0, \lambda_1, \xi_1, \lambda_2, \ldots, \lambda_n, \xi_n$ is an open chain-sequence of U, the ϵ_i and $\bar{\epsilon}_i$ are hinges of S, the relations

$$\epsilon_i p = \xi_{i-1}, \bar{\epsilon}_i p = \xi_i, \epsilon_i q = \bar{\epsilon}_i q = \lambda_i, \epsilon_i \neq \bar{\epsilon}_i$$

hold for $i = 1, 2, \ldots, n$ and $\bar{\epsilon}_i$, ϵ_{i+1} are opposed for $i = 1, 2, \ldots, n-1$. (The last condition is vacuous if $n = 1$.) The vertex ξ_0 [ξ_n] is a *tail* or *head* of (5) according as ϵ_1 [$\bar{\epsilon}_n$] belongs to $\mathfrak{r}$ or $\mathfrak{e}$ respectively. (Thus an open route-sequence of S may have two tails, two heads, or one tail and one head.) An *open route* of S is a subgraph of S derivable from an open route-sequence of S. (We shall leave the reader to guess the definitions of *subgraph of S*, *derivable* and certain other terms relating to semi-oriented graphs from corresponding definitions given for unoriented and oriented graphs.) If R is an open route of S, ξ is a vertex of R, and s is any open route-sequence from which R is derivable, then clearly $f_R(\xi) = 1$ if and only if ξ is a tail of s and $f_R(\xi) = -1$ if and only if ξ is a head of s; we shall therefore call ξ a *tail* of R if $f_R(\xi) = 1$ and a *head* of R if $f_R(\xi) = -1$. A *decomposition* of S is a set of edge-disjoint subgraphs of S whose union is S. If u, v are U-functions, a *(u, v)-decomposition* of S is a decomposition D of S into open routes such that each vertex ξ is a tail of exactly $u(\xi)$ and head of exactly $v(\xi)$ elements of D. Semi-oriented graphs are virtually a generalization of oriented graphs, since an oriented graph may be regarded as a semi-oriented graph in which each edge is incident with two opposed hinges. A *semi-orientation* of an unoriented graph U_1 is a semi-oriented graph having U_1 as its first constituent element.

Theorem 2 admits the following generalization:

THEOREM 3. *Let $S = (U, \mathfrak{r}, \mathfrak{e}, p, q)$ be a semi-oriented graph and u, v be U-functions. Then S has a (u, v)-decomposition if and only if $u + v$ is tolerable, $u - v = f$ on $V(U)$ and $(u + v) \,.\, V(H) > 0$ for each component H of U.*

The proof of Theorem 3 is a fairly easy adaptation of that of Theorem 2; but we refrained from giving the argument in this more general form to avoid obscurity. It may be remarked, however, that Theorem 1 is more readily deducible from Theorem 3 than from Theorem 2, since Lemma 8 becomes trivial if, in its statement, "an orientation" be replaced by "a semi-orientation."

Definitions. A *partition* of a set A is a set of disjoint subsets of A whose union is A. If P is a partition of $V(N)$, an N-function g is *P-tolerable* if

$$g \,.\, (S \cap T) < g \,.\, (S - T) + |S\delta|$$

for every pair S, T of subsets of $V(N)$ such that $T \in P$. A set Φ of open routes of N is *P-restricted* if no element of Φ has both its end-vertices in the same element of P.

THEOREM 2′. *Let P be a partition of $V(N)$ and u, v be N-functions. Then N has a P-restricted (u, v)-decomposition if and only if $u + v$ is P-tolerable, $u - v = f$ on $V(N)$, and $(u + v) \,.\, V(H) > 0$ for each component H of N.*

Theorem 2′ is a generalization of Theorem 2, since it clearly reduces to Theorem 2 when P is taken to be the partition of $V(N)$ into subsets of order 1. The proof of Theorem 2′, which we shall not give in detail, consists in applying Theorem 2 to an oriented graph N_1 and N_1-functions u_1, v_1 defined as follows. N_1 is obtained from N by adding, for each $T \in P$, a new vertex α_T and, for each pair ξ, T such that $\xi \in T \in P$, $u(\xi)$ new edges with tail α_T and head ξ and $v(\xi)$ new edges with tail ξ and head α_T. (Thus $|P|$ new vertices and $(u + v) \,.\, V(N)$ new edges are added altogether.) We write $u_1(\alpha_T) = u \,.\, T$, $v_1(\alpha_T) = v \,.\, T$ and $u_1 = v_1 = 0$ on $V(N)$.

Theorems 1 and 3 admit corresponding generalizations to "P-restricted" decompositions.

Since this work was a part of my thesis, I should like gratefully to acknowledge the help and guidance of my Research Supervisors, Professor D. Rees, Professor N. E. Steenrod, and Dr. S. Wylie, and the following financial support; grants from the Department of Scientific and Industrial Research, the University of Cambridge and Trinity Hall, an Amy Mary Preston Read Scholarship (awarded by the University of Cambridge), a J. S. K. Visiting Fellowship (awarded by the University of Princeton), and a Fulbright Travel Grant.

REFERENCE

1. D. König, *Theorie der endlichen und unendlichen Graphen* (Leipzig, 1936, and New York, 1950).

University of Aberdeen

HOMOTOPY AND ISOTOPY PROPERTIES OF TOPOLOGICAL SPACES

SZE-TSEN HU

1. Introduction. The most important notion in topology is that of a *homeomorphism* $f: X \to Y$ from a topological space X onto a topological space Y. If a homeomorphism $f: X \to Y$ exists, then the topological spaces X and Y are said to be *homeomorphic* (or *topologically equivalent*), in symbols,

$$X \equiv Y.$$

The relation $\equiv$ among topological spaces is obviously reflexive, symmetric, and transitive; hence it is an equivalence relation. For an arbitrary family F of topological spaces, this equivalence relation $\equiv$ divides F into disjoint equivalence classes called the *topology types* of the family F. Then, the main problem in topology is the topological classification problem formulated as follows.

The topological classification problem: Given a family F of topological spaces, find an effective enumeration of the topology types of the family F and exhibit a representative space in each of these topology types.

A number of special cases of this problem were solved long ago. For example, the family of Euclidean spaces is classified by their dimensions and the family of closed surfaces is classified by means of orientability and Euler characteristic. However, the problem is far from being solved; in fact, the topological classification of the family of three-dimensional compact manifolds still remains an outstanding unsolved problem.

To overcome the difficulty of the topological classification problem, topologists introduced weaker equivalence relations, namely, the homotopy and isotopy equivalences, which would give rise to larger but fewer classes of spaces than the topology types.

A continuous map $f: X \to Y$ is said to be a *homotopy equivalence* provided that there exists a continuous map $g: Y \to X$ such that the compositions $g \circ f$ and $f \circ g$ are homotopic to the identity maps on X and Y respectively. Two topological spaces X and Y are said to be *homotopically equivalent* (in symbol, $X \simeq Y$) if there exists a homotopy equivalence $f: X \to Y$.

It is easily verified that the relation $\simeq$ among topological spaces is reflexive, symmetric, and transitive; hence it is an equivalence relation. For any given family F of topological spaces, this equivalence relation $\simeq$ divides F into disjoint equivalence classes called the *homotopy types* of the family. Analogous to

Received November 16, 1959. This research was supported by the United States Air Force through the Air Force Office of Scientific Research and Development Command, under Contract No. AF 49(638)-179. Reproduction in whole or in part is permitted for any purpose of the United States Government.

the topological classification problem, one can formulate the ***homotopy classification*** problem in the obvious fashion.

To introduce the notion of isotopy equivalence, let us first recall the definition of an imbedding. A continuous map $f: X \to Y$ is said to be an *imbedding* provided that f is a homeomorphism of X onto the subspace $f(X)$ of Y.

A homotopy $h_t: X \to Y$, $(t \in I)$, is said to be an *isotopy* if, for each $t \in I$, h_t is an imbedding. Two imbeddings $f, g: X \to Y$ are said to be *isotopic* if there exists an isotopy $h_t: X \to Y$, $(t \in I)$, such that $h_0 = f$ and $h_1 = g$.

An imbedding $f: X \to Y$ is said to be an *isotopy equivalence* if there exists an imbedding $g: Y \to X$ such that the composite imbeddings $g \circ f$ and $f \circ g$ are isotopic to the identity imbeddings on X and Y respectively. Two topological spaces X and Y are said to be *isotopically equivalent* (in symbol, $X \cong Y$) if there exists an isotopy equivalence $f: X \to Y$.

The relation $\cong$ among topological spaces is obviously an equivalence relation. For any given family F of topological spaces, this equivalence relation $\cong$ divides F into disjoint equivalence classes called the *isotopy types* of the family. One can formulate the *isotopy classification problem* in the obvious fashion.

By the definitions given above, it is clear that every homeomorphism is an isotopy equivalence and that every isotopy equivalence is a homotopy equivalence.

Examples in the sequel will show that the converses are not always true. Hence, for any given family F of topological spaces, every topology type of F is contained in some isotopy type of F, and every isotopy type of F is contained in some homotopy type of F. Consequently, the topological classification problem can break into three steps as follows:

Step 1. *Homotopy classification*. Determine effectively all of the homotopy types of the family F.

Step 2. *Isotopy classification*. For each homotopy type α of the family F, determine effectively all of the isotopy types of the family α.

Step 3. *Topological classification*. For each isotopy type β of the family F, determine effectively all of the topology types of the family β and exhibit a representative space in each of the topology types.

In order to carry out the three steps of the topological classification problem for a given family F of topological spaces, one must make use of the various properties of spaces which are preserved by homotopy equivalences, isotopy equivalences, and homeomorphisms respectively. These properties are called the homotopy properties, the isotopy properties, and the topological properties respectively. It follows that every homotopy property is an isotopy property and that every isotopy property is a topological property. Examples in the sequel will show that the converses of these implications do not always hold.

The main purpose of the present paper is to give general tests for homotopy and isotopy properties in terms of hereditary and weakly hereditary properties with the elementary properties in general topology as illustrations. These will

be given in §§ 2 and 3. In the final section of the paper, we will describe a general method of constructing new homotopy and isotopy properties out of old ones as a striking and profound synthesis of various isolated known results.

2. Homotopy properties. A property P of topological spaces is called a *homotopy property* provided that it is preserved by all homotopy equivalences. Precisely, P is a homotopy property if and only if, for an arbitrary homotopy equivalence $f: X \to Y$, that X has P implies that Y also has P. If a homotopy property P is given in the form of a number, a set, a group, or some other similar object, P is said to be a *homotopy invariant*.

Some of the elementary properties in general topology are homotopy properties. As examples, one can easily prove the following assertions.

PROPOSITION 2.1. *Contractibility is a homotopy property of topological spaces.*

PROPOSITION 2.2. *The cardinal number of components of a topological space X is a homotopy invariant.*

COROLLARY 2.3. *Connectedness is a homotopy property of topological spaces.*

PROPOSITION 2.4. *The cardinal number of path-components of a topological space X is a homotopy invariant.*

COROLLARY 2.5. *Pathwise connectedness is a homotopy property of topological spaces.*

Nevertheless, most properties studied in general topology are not homotopy properties. To demonstrate this fact, let us first introduce the notion of weakly hereditary properties.

A property P of topological spaces is said to be *hereditary* if each subspace of a topological space with P also has P; it is said to be *weakly hereditary* if every closed subspace of a topological space with P also has P. For examples, the following properties of a topological space X are weakly hereditary:

(A) X is a T_1-space, that is, every point in X forms a closed set of X.

(B) X is a Hausdorff space.

(C) X is a regular space.

(D) X is a completely regular space.

(E) X is a discrete space, that is, every set in X is open.

(F) X is an indiscrete space, that is, the only open sets in X are the empty set $\square$ and the set X itself.

(G) X is a metrizable space.

(H) The first axiom of countability is satisfied in X, that is, the neighbourhoods of any point in X have a countable basis.

(I) The second axiom of countability is satisfied in X, that is, the open sets of X have a countable basis.

(J) X can be imbedded in a given topological space Y.

(K) For a given integer $n \geqslant 0$, $\dim X \leqslant n$. Here, the *inductive dimension* $\dim X$ is defined as follows: $\dim X = -1$ if X is empty, and $\dim X \leqslant n$ if for every point $p \in X$ and every open neighbourhood U of p there exists an open neighbourhood $V \subset U$ of p such that $\dim \partial V \leqslant n - 1$, where ∂V denotes the boundary $\bar{V} \setminus V$ of V in X (**2**, p. 153).

(L) X is a normal space.

(M) X is a compact space.

(N) X is a Lindelof space, that is, every open covering of X has a countable subcovering.

(O) X is a paracompact space.

(P) X is a locally compact space.

(Q) For a given integer $n \geqslant 0$, $\operatorname{Dim} X \leqslant n$. Here, the *covering dimension* $\operatorname{Dim} X$ is defined as follows: $\operatorname{Dim} X \leqslant n$ if every finite open covering of X has a refinement of order $\leqslant n$ (**2**, p. 153).

The first eleven properties (A)–(K) listed above are also hereditary.

A topological space X is said to be a *singleton space* if X consists of a single point. Obviously, every singleton X has all of the properties (A)–(Q). On the other hand, none of these properties prevails in all topological spaces. Hence we deduce, as a consequence of the following theorem, the fact that *none of these properties* (A)–(Q) *is a homotopy property*.

THEOREM 2.6. *Let P be a weakly hereditary topological property such that every singleton space has P and suppose that there exists a topological space X which does not have P. Then P is not a homotopy property.*

Proof. Let X be a topological space which does not have P. Consider the cone $C(X)$ over X which is the quotient space obtained by identifying the top $X \times 1$ of the cylinder $X \times I$ to a single point v, called the vertex of the cone $C(X)$. Then the space X may be identified with bottom $X \times 0$ of the cone $C(X)$ and hence X becomes a closed subspace of $C(X)$. Since P is a weakly hereditary property which X does not have, $C(X)$ cannot have P. On the other hand, it is well known that the inclusion map $i: v \subset C(X)$ is a homotopy equivalence. Since the singleton space v has P but $C(X)$ does not have P, P is not a homotopy property. This completes the proof of (2.6).

Although most of the properties studied in general topology are not homotopy properties as shown by the foregoing theorem, it is well known that almost all invariants studied in algebraic topology are homotopy invariants, namely, the homology groups, the homotopy groups, etc.

For topological spaces which are homotopically equivalent to CW-complexes, Postnikov, in his celebrated work (**3**), gave a complete system of homotopy invariants, now called the *Postnikov system* of the space. Any pair of these spaces are homotopically equivalent if and only if their Postnikov

systems are isomorphic. Hence, the homotopy classification problem of these spaces has been solved by Postnikov at least theoretically although his process is too complicated to be practicable.

3. Isotopy properties. A property P of topological spaces is called an *isotopy property* provided that it is preserved by all isotopy equivalences. Precisely, P is an isotopy property if and only if, for an arbitrary isotopy equivalence $f: X \to Y$, that X has P implies that Y also has P. If an isotopy property P is given in the form of a number, a set, a group, or some other similar object, P is said to be an *isotopy invariant*.

Most of the elementary properties in general topology are isotopy properties. For example, the eleven properties (A)–(K) listed in § 2 are isotopy properties in immediate consequence of the following theorem.

THEOREM 3.1. *Every hereditary topological property of spaces is an isotopy property.*

Proof. Let P be any hereditary topological property of spaces. Assume that $f: X \to Y$ is an isotopy equivalence and that the space X has the property P. It suffices to prove that Y also has P.

By definition of an isotopy equivalence, there exists an imbedding $g: Y \to X$ such that the composed imbeddings $g \circ f$ and $f \circ g$ are isotopic to the identity imbeddings on X and Y respectively. The image $g(Y)$ is a subspace of X. Since P is hereditary, this implies that $g(Y)$ has the property P. As an imbedding, g is a homeomorphism of Y onto $g(Y)$. Since P is a topological property and $g(Y)$ has P, it follows that Y also has P. This completes the proof of (3.1).

THEOREM 3.2. *The inductive dimension* dim X *of a topological space* X *is an isotopy invariant.*

Proof. Let $f: X \to Y$ be any given isotopy equivalence and assume that

$$\dim X = m, \dim Y = n.$$

It suffices to prove that $m = n$.

Since $\dim X \leq m$ and $f: X \to Y$ is an isotopy equivalence, it follows from the fact that the property (K) of § 2 is an isotopy property that $\dim Y \leq m$. Hence, we obtain $n \leq m$. By considering any isotopy inverse $g: Y \to X$ of f, we can also prove that $m \leq n$. Hence $m = n$ and (3.2) is proved.

Not all topological properties of spaces are isotopy properties. Examples are given by the following propositions.

PROPOSITION 3.3. *Compactness is not an isotopy property of topological spaces.*

Proof. Let Y denote the closed unit interval $I = [0, 1]$ and X the open unit interval $(0, 1)$ which is the interior of Y. It is well known that Y is compact

but X is non-compact. Hence, it suffices to prove that the inclusion $i: X \subset Y$ is an isotopy equivalence.

For this purpose, let $j: Y \to X$ denote the imbedding defined by

$$j(t) = \tfrac{1}{3}(t + 1), \qquad (0 < t < 1).$$

It remains to prove that the composed imbeddings $j \circ i$ and $i \circ j$ are isotopic to the identity imbeddings on X and Y respectively.

Define an isotopy $k_t: Y \to Y$, $(t \in I)$, by taking

$$k_t(y) = \tfrac{1}{3}(t + 3y - 2ty)$$

for each $t \in I$ and each $y \in Y = I$. Since $k_t(X) \subset X$ for each $t \in I$, k_t also defines an isotopy $h_t: X \to X$, $(t \in I)$.

Since h_0 and k_0 are the identity maps on X and Y respectively and since $h_1 = j \circ i$ and $k_1 = i \circ j$, it follows that $j \circ i$ and $i \circ j$ are isotopic the identity imbeddings. This completes the proof of (3.3).

Since the open interval $Y = (0, 1)$ is homeomorphic to the real line R, we have also proved the following corollary.

Corollary 3.4. *The unit interval $I = [0, 1]$ and the real line R are isotopically equivalent.*

Since the product of an arbitrary family of isotopy equivalences is clearly also an isotopy equivalence, we have the following generalization of (3.4).

Corollary 3.5. *For any cardinal number α, the topological powers I^α and R^α are isotopically equivalent.*

In particular, if α is a finite integer $n \to 0$, the n-cube I^n and the Euclidean n-space R^n are isotopically equivalent.

On the other hand, if α is infinite, R^α is not locally compact while I^α is compact and hence locally compact. This proves the following proposition.

Proposition 3.6. *Local compactness is not an isotopy property of topological spaces.*

4. Homotopy functors and isotopy functors. By a *covariant homotopy functor*, we mean an operator ϕ which assigns to each topological space X a topological space $\phi(X)$ and to each continuous map $f: X \to Y$ a continuous map

$$\phi(f): \phi(X) \to \phi(Y)$$

satisfying the following three conditions:

(HF1) ϕ preserves identity, that is, if f is the identity map so is $\phi(f)$.

(HF2) ϕ preserves composition, that is, if $f: X \to Y$ and $g: Y \to Z$ are continuous maps then we have

$$\phi(g \circ f) = \phi(g) \circ \phi(f).$$

(HF3) ϕ preserves homotopy, that is, if the family $h_t: X \to Y$, $(t \in I)$, of continuous maps is a homotopy, so is the family

$$\phi(h_t): \phi(X) \to \phi(Y), \qquad (t \in I).$$

If, in the preceding definition of a homotopy functor ϕ, we have

$$\phi(f): \phi(Y) \to \phi(X), \phi(g \circ f) = \phi(f) \circ \phi(g),$$

then the operator ϕ is called a ***contravariant homotopy functor***.

Similarly, by a ***covariant isotopy functor***, we mean an operator which assigns to each topological space X a topological space $\psi(X)$ and to each imbedding $f: X \to Y$ an imbedding

$$\psi(f): \psi(X) \to \psi(Y)$$

satisfying the following three conditions:

(IF1) ψ preserves identity, that is, if f is the identity imbedding so is $\psi(f)$.

(IF2) ψ preserves composition, that is, if $f: X \to Y$ and $g: Y \to Z$ are imbeddings then we have

$$\psi(g \circ f) = \psi(g) \circ \psi(f).$$

(IF3) ψ preserves isotopy, that is, if the family $k_t: X \to Y$, $(t \in I)$, of imbeddings is an isotopy, so is the family

$$\psi(k_t): \psi(X) \to \psi(Y), \qquad (t \in I).$$

One can define ***contravariant isotopy functors*** by reversing the direction of the imbeddings $\psi(f)$ and obvious modifications in (IF2) and (IF3).

Examples of homotopy and isotopy functors:

Example 1. *Topological powers.* Let n be any positive integer. Define an operator ϕ as follows. For each topological space X, let $\phi(X)$ denote the topological nth power X^n, that is, the topological product of n copies of the space X. For each continuous map $f: X \to Y$, let $\phi(f)$ stand for the nth power $f^n: X^n \to Y^n$ of f defined by

$$f^n(x_1, \ldots, x_n) = (fx_1, \ldots, fx_n).$$

Then the conditions (HF1)–(HF3) can easily be verified and hence ϕ is a covariant homotopy functor. Furthermore, if $f: X \to Y$ is an imbedding, $\phi(f) = f^n$ is clearly also an imbedding. Hence, the restriction ψ of ϕ on spaces and imbeddings is a covariant isotopy functor.

More generally, let G be a subgroup of the symmetric group S of the integers $1, \ldots, n$, that is, S is the group of all permutations of the n integers $1, \ldots, n$. Then G operates on the topological power X^n by permuting the factors of X^n. Let $\phi_G(X)$ denote the orbit space X^n/G. Since the operators in G obviously commute with the continuous maps $f^n: X^n \to Y^n$, each f^n induces a continuous map $\phi_G(f): \phi_G(X) \to \phi_G(Y)$. It follows that ϕ_G is a covariant homotopy

functor and its restriction ψ_G on spaces and imbeddings is a covariant isotopy functor.

Example 2. *Residual functors.* Let n be an integer greater than 1. Define an operator ψ as follows. For each topological space X, let $\psi(X)$ denote the residual space $X^n \setminus d(X)$ obtained by deleting the diagonal $d(X)$ from the nth power X^n. If $f: X \to Y$ is an imbedding, the nth power f^n carries $\psi(X)$ into $\psi(Y)$ and hence defines an imbedding $\psi(f): \psi(X) \to \psi(Y)$. The conditions (IF1)–(IF3) are obviously satisfied and hence ψ is a covariant isotopy functor. This isotopy functor ψ is called the nth *residnal functor* and is denoted by R_n.

Let G be a subgroup of the symmetric group of n integers $1, \ldots, n$. Then G also operates on the residual space $\psi(X)$. Let $\psi_G(X)$ denote the orbit space $\psi(X)/G$. Then $\psi(f)$ induces an imbedding $\psi_G(f): \psi_G(X) \to \psi_G(Y)$ for each imbedding $f: X \to Y$. Thus, ψ_G is also a covariant isotopy functor.

Example 3. *Mapping spaces.* Let T be a given Hausdorff space. Define an operator ϕ as follows. For each topological space X, let $\phi(X)$ stand for the space Map (T, X) of all continuous maps from T into X with the compact-open topology. For each continuous map $f: X \to Y$, let

$$\phi(f): \mathrm{Map}(T, X) \to \mathrm{Map}(T, Y)$$

denote the function defined taking

$$[\phi(f)](\xi) = f \circ \xi$$

for each $\xi: T \to X$ in $\mathrm{Map}(T, X)$. One can verify that $\phi(f)$ is a continuous map and that the conditions (HF1)–(HF3) are satisfied. Hence ϕ is a covariant homotopy functor. Furthermore, if $f: X \to Y$ is an imbedding, so is $\phi(f)$. This implies that the restriction ψ of ϕ on spaces and imbeddings is a covariant isotopy functor.

Example 4. *Enveloping functors.* Let n be any positive integer greater than 1. Define an operator ψ as follows. For each topological space X, consider as in Example 2 the nth power X^n and identify X with the diagonal $d(X)$ in X^n. Then, $\psi(X)$ stands for the subspace of $\mathrm{Map}(I, X^n)$ consisting of the continuous paths $\sigma: I \to X^n$ such that $\sigma(t) \in X$ if and only if $t = 0$. For each imbedding $f: X \to Y$, it follows from the preceding examples that the imbedding $f^n: X^n \to Y^n$ induces an imbedding of $\mathrm{Map}(I, X^n)$ into Map (I, Y^n) which carries $\psi(X)$ into $\psi(Y)$ and hence defines an imbedding

$$\psi(f): \psi(X) \to \psi(Y).$$

One can easily verify that the conditions (IF1)–(IF3) are satisfied and hence ψ is a covariant isotopy functor. This isotopy functor ψ is called the nth *enveloping functor* and is denoted by E_n. For the remaining case $n = 1$, we may define $E_1(X)$ to be the subspace of $\mathrm{Map}(I, X)$ consisting of the continuous paths $\sigma: I \to X$ such that $\sigma(t) = \sigma(0)$ if and only if $t = 0$.

For each subgroup G of the symmetric group of n integers $1, \ldots, n$, similar modifications may be made as in Examples 1 and 2.

The usefulness of these functors can be seen from the following two theorems.

THEOREM 4.1. *If ϕ is a homotopy functor, then every homotopy property of $\phi(X)$ induces a homotopy property of X.*

Proof. Let P be an arbitrary homotopy property. Assume that $f: X \to Y$ is a homotopy equivalence and that $\phi(X)$ has P. We have to prove that $\phi(Y)$ must also have P. For this purpose, it suffices to show that $\phi(f)$ is also a homotopy equivalence.

Let $g: Y \to X$ be a continuous map such that the compositions $g \circ f$ and $f \circ g$ are homotopic to the identity maps on X and Y respectively. Then there exist homotopies $h_t: X \to X$ and $k_t: Y \to Y$, $(t \in I)$, such that $h_0 = g \circ f$, $k_0 = f \circ g$, and h_1, k_1 are identity maps. By (HF3), $\phi(h_t)$ and $\phi(k_t)$ are homotopies. By (HF2), $\phi(h_0)$ and $\phi(k_0)$ are the two compositions of $\phi(f)$ and $\phi(g)$. By (HF1), $\phi(h_1)$ and $\phi(k_1)$ are the identity maps on $\phi(X)$ and $\phi(Y)$ respectively. Hence $\phi(f)$ is a homotopy equivalence. This completes the proof of (4.1).

For example, let us take P to be the pathwise connectedness. For each homotopy functor ϕ, we may define a new homotopy property which might be called the *ϕ-pathwise connectedness* as follows. A topological space X is said to be *ϕ-pathwise connected* provided that $\phi(X)$ is pathwise connected. By (4.1), we know that ϕ-pathwise connectedness is a homotopy property of topological spaces. In particular, if ϕ is the homotopy functor constructed in Example 3 with $T = S^1$ the unit 1-sphere, then one can easily see that a topological space X is ϕ-pathwise connected if and only if it is simply connected. Thus, this gives us the well-known fact that simple connectedness is a homotopy property of topological spaces.

Analogously, we have the following

THEOREM 4.2. *If ψ is an isotopy functor, then every isotopy property of $\psi(X)$ induces an isotopy property of X; in particular, every homotopy property of $\psi(X)$ induces an isotopy property of X.*

The proof of (4.2) is similar to that of (4.1) and hence omitted.

COROLLARY 4.3. *If ψ is an isotopy functor, then all homotopy invariants of $\psi(X)$, such as the homology groups of $\psi(X)$, are isotopy invariants of X.*

By suitable choices of the isotopy functors ψ, (4.3) provides many new isotopy invariants of topological spaces which enable us to solve the problems in isotopy theory. For example, let us consider a family of topological spaces

$$W_q^p, \qquad (p > 0, q > 0),$$

where W_q^p denotes the linear graph obtained by attaching p small triangles

at each end of a line-segment ab and joining the two ends of ab by q broken lines ac_kb, $k = 1, 2, \ldots, q$. Let

$$r = 2p + q.$$

Since the Euler characteristic of W_q^p is

$$\chi(W_q^p) = 1 - r,$$

it follows that the homotopy classification problem of this family of spaces $\{W_q^p: p > 0,\ q > 0\}$ is solved by the homotopy invariant $r = 2p + q$. Precisely, W_q^p and W_t^s are homotopically equivalent if and only if

$$2p + q = 2s + t.$$

For the isotopy classification of the spaces W_q^p with the same $r = 2p + q$, let us use the second residual functor R_2. In **(1)**, it has been computed that the two-dimensional homology group of $R_2(W_q^p)$ is a free abelian group of rank $2p^2$ and the one-dimensional homology group of $R_2(W_q^p)$ is a free abelian group of rank $6p^2 + 4pq + q^2 + 2p + q - 1$ for all W_q^p with $2p + q > 0$. This solves the isotopy classification problem for the spaces W_q^p.

References

1. S. T. Hu, *Isotopy invariants of topological spaces*, Technical Note, AFORS TN59-236, AD212006. Also to appear in the Proceedings of the Royal Society, England.
2. W. Hurewicz and H. Wallman, *Dimension Theory* (Princeton Mathematical Series, No. 4 [Princeton University Press, 1941]).
3. M. M. Postnikov, *Investigations in homotopy theory of continuous mappings I, II*, Trudy Mat. Inst. Steklov, No. 46, Izdat. Akad. Nauk SSSR (Moskow, 1955). Amer. Math. Soc. Translations, Series 2, 7 (1957).
4. A. Shapiro, *Obstructions to the imbedding of a complex in a euclidean space I*, Ann. Math., *66* (1957), 256–269.
5. W. T. Wu, *On the realization of complexes in Euclidean spaces I-III*, Acta Math. Sinica, *5* (1955), 505–552; *7* (1957), 79–101; *8* (1958), 79–94.
6. ——— *On the imbedding of polyhedra in Euclidean spaces*, Bull. Polon. Sci. Cl. III, *4* (1956), 573–577.

University of California at Los Angeles.